Self-Learning Material (SLM) for the Course of

M.A. / M.Sc. in Geography

Directorate of Open and Distance Learning (DODL)

University of Kalyani

Paper: GEO309T

(Total Credit – 4; Total Marks – 100: Internal Evaluation – 20 + Semester-end Examination - 80)

University of Kalyani

Kalyani

Nadia – 741235

West Bengal

Course Materials Compiled by:	
Dr. Sayan Choudhary Assistant Professor (Contractual) Directorate of Open and Distance Learning University of Kalyani, West Bengal	**Group – GEO309T.1: Geographical Thought**
Dr. Bhaswati Mondal Assistant Professor (Contractual) Directorate of Open and Distance Learning University of Kalyani, West Bengal	**Group – GEO309T.2: Economic Geography, Transport Geography and Geography of Trade**

December 2019

Directorate of Open and Distance Learning, University of Kalyani.

Published by the Directorate of Open and Distance Learning,

University of Kalyani, Kalyani-741235, West Bengal.

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2017 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor (Dr.) Sankar Kumar Ghosh, Hon'ble Vice- Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks is also due to the Course Writers-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and co-ordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self writing and collected from e-book, journals and websites.

Director

Directorate of Open and Distance Learning

University of Kalyani

Syllabus

Paper – GEO309T: (Total Credit - 4, Total Marks – 100)

Group – GEO309T.1: Geographical Thought
(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit. 01	General character and development of Geographic knowledge: Contributions of Greek, Roman and Indian scholars during the ancient period and Arab scholars during the medieval period;
Unit. 02	Different schools of Modern Geography;
Unit. 03	Concept of paradigm shifts in Geography;
Unit. 04	Dualism and dichotomies in Geography: Physical Geography and Human Geography, Regional Geography and Systematic Geography, Ideographic approach and Nomothetic approach;
Unit. 05	Positivism and Quantitative revolution in Geography; Hartshorne-Schaefer debate, System approach in Geography;
Unit. 06	Critical revolution in Geography; Humanistic Geography; Radical Geography; Behavioural perspectives in Geography;
Unit. 07	Welfare Geography; Feminism and Feminist Geography;
Unit. 08	Postmodernism and Postmodern Geography

Group – B: Economic Geography, Transport Geography and Geography of Trade

(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit. 09	Scope and Advancement of Economic Geography; Concept and Classification of Resources;
Unit. 10	Concept of Agricultural Region; Concept and Measurement of Agricultural Productivity and Efficiency; Green Revolution and White Revolution in India
Unit. 11	Concept of Industrial Region and Industrial Complex; Special Economic Zone; Growth of IT Industry in India; Industrial Policy of India: Role of Liberalization; Privatization and Globalization
Unit. 12	Concept of Digital Divide
Unit. 13	Concepts and Measures of Distance, Accessibility and Connectivity; Transport Cost: Factors and Comparative Cost Advantages; Freight Corridor

Unit. 14	Concept of Ring road, By-pass, Golden Quadrilateral, North-South and East-West Corridor; Significance of Trade in Regional and National Economy
Unit. 15	Concept of Export Processing Zones, Exclusive Economic Zones, Forward Trading and E-commerce; Freight Equalization Policy on Indian Trade
Unit. 16	Role of GATT and WTO in International Trade; Issues Related to FDI in India's Retail Sector and Cashless Economy

Contents

Paper – GEO309T: (Total Credit - 4, Total Marks – 100)

Group – GEO309T.1: Geographical Thought
(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

1.1 INTRODUCTION

Geography as a discipline evolved through different stages of historical development of mankind. During its inception geography was chiefly confined with descriptive studies of man and nature. The Greek, Roman and Indian scholars during ancient time made valuable contributions in the development of geography. Medieval period was chiefly a dark period for over all development of science. Modern geography was established by pioneers in Germany, France, Britain, Russia and the United States. During this time dualism and dichotomies in geography became prominent. Positivism and quantitative revolution in geography opened new vista in geographic research. Dissatisfaction with quantitative revolution gave way to critical revolution, welfare geography and feminist geography.

1.2 LEARNING OBJECTIVES

The present section aims to introduce the following topics -

- General character and development of Geographic knowledge: Contributions of Greek, Roman and Indian scholars during the ancient period and Arab scholars during the medieval period;
- Different schools of Modern Geography;
- Concept of paradigm shifts in Geography;
- Dualism and dichotomies in Geography: Physical Geography and Human Geography, Regional Geography and Systematic Geography, Ideographic approach and Nomothetic approach;
- Positivism and Quantitative revolution in Geography; Hartshorne-Schaefer debate, System approach in Geography;
- Critical revolution in Geography; Humanistic Geography; Radical Geography; Behavioural perspectives in Geography;
- Welfare Geography; Feminism and Feminist Geography;
- Postmodernism and Postmodern Geography

1.3 ASSESSMENT OF PRIOR KNOWLEDGE

Discussion about nature, scope and relevance of geography is necessary. Discussion about different branches of geography is necessary.

1.4 LEARNING ACTIVITIES

Preparation of short notes and essays on different topics of discussed.

1.5 FEEDBACK OF LEARNING ACTIVITIES

Debate and discussion on various topics discussed in the paper may be conducted.

Class seminar on various topics may be arranged

1.6 EXAMPLES AND ILLUSTRATIONS

UNIT - 1: GENERAL CHARACTER AND DEVELOPMENT OF GEOGRAPHIC KNOWLEDGE: CONTRIBUTIONS OF GREEK, ROMAN AND INDIAN SCHOLARS DURING THE ANCIENT PERIOD AND ARAB SCHOLARS DURING THE MEDIEVAL PERIOD

In its wider connotation as a branch of knowledge concerned with the satisfaction of human curiosity about the lands and peoples away from one's home base, speculation regarding mysteries of the physical environment, and the role it plays in shaping the destiny of man upon the earth, geography is as old as human civilization. As such, each major cultural realm has had its own historiography of geography. Modern geography as practised over the entire world today represents, however, an outgrowth of the European geographic tradition so that the historiography of modem geography is essentially an account of the conceptual developments among Europeans regarding the nature of the earth and its environment and the way it influences man. Thus, the roots of modem geography are to be traced back to the thought of the ancient Greeks.

CONTRIBUTIONS OF THE GREEKS AND THE ROMANS

References to descriptive writings about lands and peoples in different parts of the earth's surface are found in the oral traditions of classical Greece and are reflected in the writings of Homer, whom the Greek geographers had themselves referred to as the father of geography. *Odyssey,* Homer's epic poem written sometimes in the ninth century B.C., presents geographical accounts of the lands and peoples located on the margins of the world then known to the Greeks. The poem records the wanderings of Odysseus to return to Ithaca after the fall of Troy, when he was blown off course by a storm, and it took him twenty years to reach home. The poem contains a geographical account of the distant places visited by the hero of the epic in course of his long journey. In it there are references to a land of continuous sunshine, and later of Odysseus's visit to an area of continuous darkness. Apparently a Greek poet could not have imagined these scenes. Somehow accounts about the nature of the earth in the far north of Europe during the long summer days and the continuous winter darkness had filtered back to Greece, and were woven with other geographical threads into an enchanting adventure story. As in the case of *Meghdoot* of Kalidas, many have tried to identify the many places referred to in the epic poem.

However, Thales of Miletus-a town located near the mouth of the Menderes river on the eastern side of the Aegean Sea (which was both a centre of learning and a flourishing centre of commerce)-who lived in the seventh and the sixth century B.C., is regarded as the first Greek to have devoted focused attention to the measurement and location of places on the surface of the earth. Thales himself was a very successful businessman. In the course of a business trip to Egypt, he had been greatly impressed by the geometrical traditions of the Egyptians in the measurement and computation of areas. He had introduced some of these ideas among the Greeks. Anaximander, a contemporary of Thales and a few years his junior, is credited with having first introduced the idea of the sundial consisting of a pole set vertically over a flat surface to measure the varying position of the sun by measuring the length and direction of the shadow cast by the pole. The shadow was shortest at noon and provided an exact north-south line for determining the correct longitude of the relevant place. Anaximander is also said to have produced a map of the known world with Greece as its centre. Thales and Anaximander have jointly been regarded as the originators of the mathematical tradition in geography in ancient Greece. The literary tradition in the writing of geography had also developed around the same time. Hecataeus, a resident of the town of Miletus, and born around 475 B.C.-about the time that Thales and Anaximander had passed away·-originated the literary tradition and his book *Periods Ges* (Description of the Earth) is regarded as the first known attempt to synthesize available knowledge about the world in a usable form. Hecataeus is also one of the earliest writers of prose in classical Greek literature.

The next great name in this context is that of Herodotus (circa 485- 425 B.C.) who is widely known as the father of history, but is also generally regarded as one of the founders of geography. His history of the Greek struggle with the "barbarians" included (as digressions) descriptions of various places visited by the author. Herodotus firmly believed that all history must be treated geographically and also that all geography must be studied historically. For him geography provided the stage, or the setting that gives meaning to historical events. Herodotus had travelled a great deal. Throughout his travels he had retained a keen interest in the nature of the landscape so that he not only described geographical phenomena but also tried to explain them. Examples include his attempt to explain the annual fluctuations in the flow of the Nile, and the processes involved in the origins of deltas occurring at the mouth of the Meander (Menderes) river at Miletus. Herodotus had no interest in the mathematical tradition and showed no interest in problems like measurement of the earth's circumference. He accepted the Homeric view of the earth as a flat disc over which the sun was believed to travel in an arc from east to west.

Plato (428-348 B.C.) also made an important contribution to the development of geographical ideas. Plato was a great proponent of deductive reasoning. He insisted that the observable phenomena on the earth's surface represent poor copies of ideas from which these observable phenomena had degenerated. By way of illustration he referred to the case of Attica (the ancient territory of which Athens was the capital). According to Plato, the area was originally very fertile and capable of supporting a large population of men and animals. He wrote that compared to its original state, the Attica of his time was "like the skeleton of a sick man, all the flesh and soft earth having been wasted away, and only bare framework of the land being left" (cited in Glacken, 1967, p. 121). Contemporary philosophers in Greece generally accepted the idea that symmetry of fotn, is one of the essential alttibutes of perfection, and that the most completely symmetrical for11, was a sphere. It was argued that since the earth had been created to serve as the home of man, it must have a perfect form, and therefore it must be a sphere. Plato is regarded as the first scholar who put forward the concept of a spherical earth located in the centre of the universe, and the sun and all the other celestial bodies moving around it. Plato offered no argument or evidence as proof that the earth is round. Providing the proof for the spherical shape of the earth was left to Aristotle (384-322 B.C.), who was a student and a member of Plato's academy for twenty years.

Aristotle is regarded as the pioneer of inductive reasoning and the inductive approach to acquiring knowledge. He was convinced that the best method of building a reliable theory was to start with the observation of facts. This required reasoning from the particular to the general, in contrast to Plato's deductive approach which required the student to proceed from the general to the particular. Aristotle laid the foundation of what has been regarded as the world's first paradigm to guide research procedures. He laid down four fundamental principles of scientific explanation: First, it is necessary to establish the necessary characteristics (i.e., the nature) of the phenomenon being investigated; Second, it is necessary to identify the substance of which it is composed; Third, it is necessary to identify the process through which the phenomenon has attained its present form; and lastly, it is necessary to identify the purpose that the phenomenon concerned fulfils in the overall scheme of nature. This last principle makes Aristotle stand out as the first teleologist in that he believed that everything was changing in accordance with a preexisting plan.

Aristotle argued his propositions so convincingly that his research methodology appeared irrefutable at the time it was presented. His intellectual status in the contemporary world of scholarship was so high that his ideas were accepted without question for a long time. (Some of his ideas were patently false, however. One such was the idea that habitability on the earth surface is a function of distance from the equator, and that areas around the equator are too hot for human survival.)

Although Plato and Aristotle gave intellectual leads that contributed to the development of knowledge about the earth as human habitat, neither of the two could be identified as a geographer. As contrasted to this, Erastosthenes (276-194 B.C.) is often referred to as the father of geography as a branch of knowledge. He is said to have coined the word geography. The term is derived from ***ge,*** meaning the earth, and ***graphe,*** meaning description. Thus was born geography as a field of study which specialized in presenting reasoned description of the Greek ecumene, and speculated about the nature of peoples and places beyond the range of knowledge in contemporary Greece. He wrote the first formal text on geography entitled ***Geographica.*** His estimate of the earth's circumference was remarkably accurate, and, therefore had proved a major step forward in the development of knowledge about the earth. Erastosthenes was the chief librarian at the famous museum at Alexandria-a post that he occupied for about forty years until his death in 194 B.C. Under his leadership, the museum had developed into a major centre of astronomical research, a field of knowledge that was at that time viewed as closely associated with geography. Erastosthenes identified five climatic zones, one torrid zone, two temperate zones, and two frigid zones. He also improved upon the Aristotelian idea on this subject by giving latitudinal boundaries to the five climatic zones. The ***torrid zone*** extended 24° north and south of the equator, and the ***frigid zones*** extended to 24° from either pole. The areas in between were the two ***temperate zones.***

After the death of Erastosthenes the post of chief librarian of the museum at Alexandria went to Hipparchus who was the first to divide the circle into 360 degrees. He also defined a grid of latitudes and longitudes for the earth, and identified the equator as a great circle that divides our spherical planet into two equal parts. Hipparchus pointed out that since the earth makes one complete revolution in 24 hours, it covers a journey of 360 degrees in a day and so covers fifteen degrees of longitude in one hour. He also made a significant contribution to the development of map projections by suggesting ways for overcoming the difficulty of representing the spherical earth on a flat sheet of paper.

The cartographical cosmographical traditions set by Erastosthenes and Hipparchus were further advanced by the succeeding generation of students at the museum. The cumulative knowledge gained through these exertions culminated in Ptolemy's (9-168 A.D.) eight volume work entitled ***Guide to Geography.*** Ptolemy was himself a great astronomer of his time and was the author of the famous text on classical astronomy entitled ***Almagest*** which had for long remained the most standard reference on the movement of celestial bodies. His ***Guide to Geography*** was also of related interest. By adopting the system of latitudes and longitudes based on the division of the circle into 360 degrees, he attempted to give precise location for all the known places in precise mathematical te1111s. Six out of the eight volumes of his ***Guide to Geography*** consisted of tables of latitudes and longitudes. The first volume was devoted to a discussion on map projections, and the eighth volume contained maps of different parts of the world showing all the places that had been included in volumes two to seven. It is true **that** from the perspective of the present, Ptolemy's book would appear as **a** monumental collection of errors. It was, however, a piece of great scholarship at the time when it was originally presented. Ptolemy's calculations of latitudes and longitudes are found to be wrong since these calculations had been based on estimated lengths of journeys between places; and these could never be accurate. Another major source of error was that Ptolemy had rejected Erastosthenes' almost correct estimate of the earth's circumference in favour of Posidonius's (which gave a figure that fell short of the actual by a little over one-fourth). (Erastosthenes had estimated the earth's circumference at 252,000 stadii, and Posidonius at 180,000 stadii (one stadium being equal to 157.5 metres.)

Strabo (64 B.C. to 20 A.D.), born a century-and-a-half before Ptolemy, had carried forward the tradition of topographical work of Greek geography as started in the works of Herodotus. His seventeen-volume work named *Geography* was largely an encyclopaedic description of the world known to the Greeks. Unlike the works of most other Greek scholars, Strabo's book was found almost intact. The first two volumes of his book contain a review of the work of other geographers since the time of Homer. They give a fair idea of the nature of geographical writing in ancient Greece. The next eight volumes were devoted to Europe, six to Asia, and one to Africa. Strabo's book was written to cater to the needs of a specific group of readers, namely the officers of the administration, statesmen, and commanders of the Roman empire. The purpose was to provide a handbook of information about places and people to help the imperial officers in the better appreciation and accomplishment of their task.

Strabo's book had laid down a clear foundation for chorological writing in geography. Explaining the method of writing geography, Strabo wrote:

... just as the man who measures the earth gets his principles from the astronomer and the astronomer his from the physicist, so too, the geographer must in the same way take his own point of departure from the man who has measured the earth as a whole, having confidence in him and in those in whom he, in his tum, had confidence, and then explain in the first instance, our inhabited world, its size, shape and character, and its relation to earth as a whole; for this is the particular task of the geographer. Then,

secondly, he must discuss in a fitting manner the several parts of the inhabited world, both land and sea, noting in passing where the subject has been treated inadequately by those of our predecessors whom we have believed to be the best authorities on the matters (Strabo, trans., Jones, 1917, pp. 429-431; cited in James, 1972, p. 47).

Both Strabo and Ptolemy had lived at a time when the Roman empire was at its zenith. It was the largest centralized empire in history till that time. The state needed to have exact description of its territories as well as the other territories it interacted with. This knowledge was necessary both for effective administration and trade as also for the training of the younger generation from among whom the future crop of administrators was to be recruited. The work of the two Greek scholars, besides extending the frontiers of knowledge, was designed to meet a definite need of society. Geography was flourishing because it served a useful purpose.

GEOGRAPHY IN THE MIDDLE AGES

During the fifth century A.D., the Roman empire suffered demise. The central administration had greatly weakened and consequently the constituent territories gradually became independent. As trade and commerce declined, the geographic horizons of the people rapidly narrowed down so that, with the passage of time, the geographical horizon of most people in Christiandom became confined to their immediate surroundings. Given the extremely narrow world-view of contemporary European societies, it was natural that religious orthodoxy should increase. Before long, scriptures had begun to be regarded as the ultimate repository of knowledge of every kind so that an impression was created that there was no need to learn anything outside the Holy books. Anything that did not conform with the "truth" of the scriptures was regarded as the product of a perverse mind and had, accordingly, to be rejected. Under these conditions, science (and, therefore, geography) could not develop and the Middle Ages represented the Dark Age in the history of scientific knowledge in Europe. During this long period, scientific concepts developed by the ancient Greeks were reshaped with a view to make them conform with the "truth" preached by the Church. For example, the idea of a spherical earth was abandoned in favour of the old concept of the earth as a flat disc, with Jerusalem as its centre. This dismal state of affairs continued almost until the end of the twelfth century A.D.

By the end of the eleventh century A.D., overland travel of Christian pilgrims to Jerusalem across Turkey and Syria had been made very difficult on account of Muslim domination over these territories. This aroused the religious sentiments of Christian Europe. A series of military campaigns were organized with a view to rescue the Holy Land of Jerusalem from the control of Muslims. Between 1096 and 1270 A.D., eight different crusades were organized for the purpose. These crusades (religious wars) played a major role in broadening the geographical horizon of Christian Europe. Men from different parts of Europe had come together to participate in them. These participants went back to their homes with new knowledge and information about the landscapes and customs of many areas beyond the range of the familiar. This stimulated interest in, and the urge to gain knowledge about, unfamiliar places. The religious wars, therefore, had led to a new beginning-a revival of interest in geography as a branch of knowledge. Expeditions began to be organized to distant places. The most famous of such expeditions was Marco Polo's voyage to China, the Far East, and the Indian Ocean undertaken between 1271 and 1295.

The Crusades proved a stimulant to the revival of interest in the study of peoples and places in far-off lands in another way also. Owing to the "religious" wars the Muslims had closed the overland routes to India and beyond to European merchants who had until then participated in the highly profitable spice trade between India and Europe. Attempts were, therefore, directed to finding an alternative route to the Indies. Two such attempts led to the glorious discoveries of Columbus and Vasco de Gama.

GEOGRAPHY IN THE ARAB LANDS

The fall of the Roman empire, and the decline of scientific learning in Europe was followed by a period of great ascendancy in the Muslim world which, under the influence of Prophet Mohammed, had been transformed from a multitude of tribes divided by intertribal feuds into integral components of a larger all-inclusive identity based on adherence to a common set of religious beliefs and practices. The followers

of Islam soon embarked upon a course of conquest of the world outside Arabia with a view to spreading the new religious ideology to the farthest corners of the world. Persia and Egypt were conquered in 641-642 A.D., and by A.D. 732 the whole of the West Asian desert region was under their control. They soon overran the Iberian peninsula and Spain and Portugal remained under Muslim rule for almost nine hundred years. Muslim influence also extended eastward into India and parts of south-east Asia. The act of holding on to such a huge politico-cultural empire had, in itself, become a major stimulant to the rise of interest in geographical learning. The Arabs held a monopoly
over the spice trade between India and Europe. This trade required a great deal of travel over land and sea. Travels between places spread over such a large expanse of territory became the source of considerable extension of knowledge about geographical environment in tropical regions.

Following the widening of its geographical horizon, the Arab world became fired with a new zeal for scientific learning. Baghdad (founded in 726 A.D.) became a major centre of learning. Its rulers (the Caliphs) employed learned men of different faiths to make authentic translations of the major scientific works of their respective languages. Included in these works were books on astronomy and geography. Scholars were also employed to calculate the circumference of the earth, and to fix latitudes in the plain of the Euphrates. The method employed was the one used by Erastosthenes about a thousand years earlier. Available texts on geography written by the ancient Greeks (including Ptolemy) were translated into Arabic, and new texts were got written after duly incorporating the new knowledge derived from the records of observations made by Arab merchants and explorers.

Thus, as a result of Ibn-Hakul's voyage to the south of the equator (made between 943 to 973 A.D.), the wrong notion regarding the inhabitability of the torrid zone (as perpetuated by Aristotle) was abandoned. Around the same time, in course of his travels down the east coast of Africa upto the Mozambique, Al-Masudi had reported the phenomenon of monsoonal winds. Another contemporary named Al-Maqdisi had established (in 985) the general truth that the climate of any place is a function not only of its latitude but also of its position on the east or west side of a landmass. He is also credited with the knowledge that most of the earth's landmass lies north of the equator. Al-Idrisi (1099-1180 A.D.) made many corrections to Ptolemy's book. On the basis of the new information collected by Arab explorers, and some collected from other sources, he produced his own book on geography (in 1154) in which the Greek idea about the Indian Ocean as a closed sea was corrected. Also corrected were the positions of many rivers including the Danube and the Niger.

Another great Arab explorer was Ibn Batuta (1304-1368 A.D.). He extensively explored regions of North Africa and West Asia, sailed along the Red Sea, and travelled south along the east coast of Africa to Kilwa, a point about 10 degrees south of the equator. He had reported that an Arab trading post was located in the latitude of 20 degrees south thereby confirming the habitability of the torrid zone. Ibn Batu ta had also travelled overland from Mecca to Persia, Bukhara and Samarkand, and from there across Afghanistan to Delhi. I-le had visited several islands, including Ceylon (Sri Lanka), Sumatra, and the Maldives. He also visited China and returned to Fez, the capital of Morocco, in 1350 A.D. From there he travelled across the Sahara to Timbuktu on the Niger, gathering valuable information about Black Muslims on the way. He settled down in Fez in 1353 A.D. after travelling an estimated 75,000 miles, a world record for his time. On the request of the king of Morocco, he put down a detailed account of his travels for posterity. The last great Muslim scholar who contributed significantly to the development of geographical knowledge was Ibn-Khaldun (1342-1405 A.D.) who wrote (in 1377) a detailed introduction to world history published under the title *Muqaddimah.* In his introduction to the book he identified two sets of influences on man's progress (i.e., history): One, the physical environment, and two, the social environment derived from culture and belief rather than the natural environment. This distinction between the two sets of environmental influences on man was a remarkable intellectual achievement for his time so that Kimble (1938) was prompted to remark that Ibn-Khaldun had "discovered ... the true scope and nature of geographical inquiry".

lbn-Khaldun had concentrated on the study of the tribe and the city the two most important elements in the political organization of the desert society in the Arab world. He identified the tribe and the city as two distinct stages in the evolution of social organization in a desert environment. While the nomads represented an earlier (primitive) stage of social organization, the city dwellers represented the last stage in the development of social life, almost the point where decay sets into the social organization owing to the sedentary lifestyle of the urban community. Many credit Ibn Khaldun with having presented in this way one of the earliest concepts of the life cycle of the states. Surprisingly, however, Ibn-Khaldun had clung to the Aristotelian idea about the inhabitability of the equatorial regions. To the great credit of lbn-Khaldun is the fact that he was the first great scholar to direct attention specifically to the study of the man-environment relationships.

The significance of Arab contribution to the historiography of modern geography lies in that the development of geographical knowledge in the Arab world represented, in some ways, a further development over the original base provided by the geographers of ancient Greece, whose works had been translated into Arabic, and widely used by Arab scholars. Thus, while Europe itself had forgotten the Greek heritage in geography, the Arabs had held the banner aloft and it was largely through contact with the Arab world, and the translation of their books (including retranslation of Greek works from the Arabic translations) that geography got revived as a living science in fifteenth century Europe. Indeed, the countries having the closest contacts with the Arab world, such as Spain and Portugal, spearheaded the series of exploratory voyages that ultimately opened the way to the revival of interest in geographical learning. Large parts of Spain and Portugal had remained under Muslim rule since the eighth century A.D. While Portugal had become free in the middle of the thirteenth century, the Spaniards pushed out the Muslims gradually from the peninsula through a series of efforts lasting over a century from 1391 to 1492.

Both the Portuguese and the Spaniards had mastered the art of shipbuilding and navigation, and had launched ambitious programmes of voyages of exploration with a view to promoting trade and commerce with the outside world, particularly the spice trade with the Indies, and trade in gems and precious stones with parts of Africa south of the Sahara. Since the overland routes in each case were then under the control of the Arabs, it was necessary to find alternative sea routes, which contemporary science had shown to be well within the pale of possibility in view of the round shape of the earth, and the continuity of the oceans. The importance of the spice trade for contemporary Europe lay in that owing to the inadequate supply of sugar, spices were required to make food palatable. Besides, in the absence of refrigeration, meat was to be stored in dried and salted form. Such meat required spices in order to be made reasonably palatable. What was more, Genoa and Venice which had earlier been flourishing centres of trade in spices from India and beyond, were now deserted as the Arabs had blocked direct contact between Europe and the regions of supply further east in Asia.

THE AGE OF EXPLORATION

Portugal's Prince Henry "The Navigator" who had (in 1415) succeeded in capturing the Muslim base at Ceuta on the southern side of the Strait of Gibraltar, took the first initiative toward wider exploration across the high seas. From his Muslim prisoners, the Prince had learnt that many of the most valuable items of merchandise traded in European markets by the Arabs were brought from areas in Africa to the south of the Sahara. This inspired him to sponsor sea voyages of exploration along the western coast of Africa. Around this time-in 1410-two important publications had appeared in geography. One was the Latin translation of Ptolemy's *Geography* made from a copy preserved in Byzantium (Istanbul), and the other was a book called *Imagine Mundi* authored by Pierre d' Ally in which he presented a summary of various geographical writings then existing in the countries of Christian Europe. The two were very influential in promoting interest in geographical knowledge and created a favourable climate for the launching of voyages of exploration, and for developing better techniques of cartography and map design.

Prince Henry was responsible for establishing a broad-based institute of geographic research at Sagre near the port of Lagos where a rich library of all the available literature in geography, cartography, astronomy and related subjects was stocked, and scholars (including linguists) from all parts of Europe were invited to teach Portuguese students the art of navigation, and to inform them about the existing knowledge regarding the earth and its environment in different parts of the world so that suitable preparations could be made to meet the contingent situations likely to be encountered in the process of exploratory voyages then being planned for the exploration of the western coast of Africa, and to find an alternative route to the spice islands beyond India. These explorations initiated by the Prince laid firm foundations for the larger ventures by subsequent explorers under the patronage of the royal house of Portugal, culminating in the great voyage of Christopher Columbus (who discovered the new World in 1492) and the discovery of an alternative sea route to India by Vasco da Gama in 1498.

Columbus had studied at Sagre, and he had been greatly influenced by Pierre d'Ally's *Imagine Mundi* which had suggested that since the earth was round, a route to China and India could be found by sailing west from the Canary Islands. It is a different matter though that in the process of finding an alternative route to the Indies, Columbus landed in America rather than Asia, his intended destination. Columbus died in 1506 still believing that he had discovered a part of Asia. The task of finding an alternative route to Asia by sailing west and then north along the coast of South America was accomplished by another great Portuguese explorer three decades later in October 1520. This voyage was performed by Magellan whose name the (Magellan) Strait connecting the Atlantic with the Pacific now bears.

Voyages across the sea required maps and charts to guide the sailors in course of their travels. Ptolemy's map was used in the beginning. It was the task of the royal cartographers to correct the old map in the light of new information then available. Thus, Venice and Genoa soon emerged as great centres of cartographic learning. European sailors and merchants departed from either of the two port towns for their journeys to the eastern Mediterranean to pick up the cargo brought by Arab ships from the East. The first globe showing the earth as a sphere was produced by Martin Behain in 1490, and map projections tackling the problem of representing the round earth on a flat sheet of paper began to receive attention of scholars soon after. In 1530, Peter Apian produced a heart-shaped map of the earth in which both latitudes as well as longitudes were shown as curved lines. Neither distance nor direction was represented correctly, and the map showed only one hemisphere. Apian's student, Gerhard Kramer (who later adopted the name Gerhard Mercator) made a world map in 1538 by joining two heart-shaped projections:-one for each hemisphere. Mercator earned celebrity in 1569 when he succeeded in designing a projection that showed the whole of the earth surface on a single network of latitudes and longitudes. This was the famous Mercator Projection-the orthomorphic cylindrical projection. As we know, even though theoretically an orthomorphic projection, it greatly distorted the shapes of continents, but its great advantage lay in that on it compass bearings could be shown by straight lines so that navigators could plot their course without being required to draw cumbersome curves. The projection could not be easily used until English geographer Edward Wright (1558-1615) produced the trigonometric table to reproduce the projection. This improvement made the Mercator projection universally accepted for maps on which to base navigators' charts. Focus on improvements in cartographic techniques continued through the sixteenth and seventeenth centuries. New projections were devised and old map projections improved upon. Map makers remained busy revising old maps in the light of new information obtained from travellers and explorers. From the time of Magellan (who explored the outlines of South America between 1518 to 1521), and James Cook (1728-1779) (who through his three different voyages, performed between 1768 to 1779, drew the outlines of the Pacific Ocean and eliminated the possibility of the existence of Ptolemy's Southland), scholars were directly addressing the task of drawing correct outlines of landmasses and water bodies. They were also busy devising techniques of surveying and cartography to be able to present true-to-scale reality of the earth's surface on their maps and charts. This task was almost complete by the time of James Cook's death in 1779 and a good deal of new information about world climates, wind regimes, distribution of flora and fauna, and patterns of human civilization over the earth surface had been obtained. Incorporating the ever increasing information and data with a view to presenting a correct and meaningful description of the earth surface had become a formidable task. The challenge posed by the problem attracted a number of leading scientists to the study of geography.

While explorers were busy fixing the outlines of continents and oceans, and cartographers remained busy in drawing more accurate representations of the earth surface on maps, the world was experiencing a great revolution in knowledge about the nature of the universe and the earth's position in it. The old-time concept of the earth as the centre of the universe was abandoned in favour of the concept of a heliocentric universe first put forward by the Polish scholar Nicolaus Copernicus in 1543. The concept was further refined by Kepler (1571-1630) in 1618 and Galileo (1564-1642) in 1623. Galileo further revolutionized scientific thinking by formulating the concept of mathematical order in the universe i.e., an order in which relationships between phenomena could be described in terms of mathematical laws rather than verbal logic. A further scientific advance came in the form of Newton's law of gravitation in 1686. Thus, in the course of a century and a half, seeds of scientific revolution had been sown. These heralded the beginning of the rise of specialized branches of knowledge, each focusing on some particular theme, object, or relationships between phenomena. The rise of specialized systematic sciences, each focusing on a particular category of facts or relationships signalled the demise of the era of universal scholarship and of cosmographies in which scholars had attempted to bring together all that was known about the earth and its parts in single volume works in the style set by Strabo.

In view of the rapid flow of new information derived from the increasing stream of explorations and scientific research, compilation and synthesis of knowledge in a meaningful manner became an increasingly challenging task that required a high degree of scholarship. The cosmographers of that period were, therefore, far from mere popularizers. The first great cosmographer of the age of exploration was the German scholar Sebastian Munster (1489-1552) who had been engaged, with the help of 120 other authors and artists, in writing a broad-based cosmography incorporating the latest information on every important aspect for over eighteen years. The outcome was a six-volume work entitled *Cosmographie Universal is* published in 1544. Written in the tradition of Strabo's *Geography,* the book earned its author the popular title of "the German Strabo". The first volume of Munster's cosmography presented a general picture of the earth on the lines of Ptolemy's *Geography,* while the remaining five volumes were devoted to descriptive accounts of the major divisions of the earth's surface.

Munster's work was a combination of tradition (imaginative stories about people and places which were part of popular belief) and science (incorporating new information derived from explorations and scientific investigations). Thus, his account of America and Africa included stories of men with heads on their chests, and having a combined animal and human form. Such beliefs were part of contemporary scholarship and consequently, Munster's volumes were avidly read necessitating several editions between 1544 and 1550; and the book remained a popular reference for about a century thereafter. Another leading cosmographer was the German scholar Cluverius (1580-1622) who had published a six-volume compendium on universal geography (following the general plan of Munster's work but better informed) in 1624. The first universal geography to appear in the English language was written by Nathanael Carpenter (1589-1628) a scholar at Oxford, who had benefited from his association with Cluverius during the latter's frequent visits to Oxford. Carpenter's book had appeared in 1625, a year after the publication of Cluverius's book.

From Cosmography to Scientific Geography: Contribution of Bernard Varenius

In course of time, the tradition of writing cosmographies got concretized into a coherent body of knowledge that came to be described as "general geography". *Geographia Generalis* (1650) of Bernard Varenius (1622-1650) (a Dutch scholar) was an outgrowth from the cosmographic tradition even though it is rightly regarded as a major step forward toward laying the foundation of scientific geography. Varenius' s book was, according to Dickinson (1969), the first work "which sought to combine general, mathematical, and physical geography and chorology". Varenius set forth clearly the distinction between two forms of geographical scholarship-the one concerned with the description of particular places (i.e., regional description), and the other concerned with developing general laws and hypotheses of wider applicability. He termed the first Special or Particular Geography (i.e., geography of particular places etc.) and the second as General Geography.

Varenius was writing at a time when voyages of exploration were pouring in a flood of new information and data so that one of the major problems facing contemporary scholars was how to relate specific pieces of information to general principles. Among geographers, Varenius was the first to focus attention on this problem; and the solution that he offered through his *Geographia Generalis* was to become the basic tenet of geography as a branch of knowledge which has ever since retained a twofold division into Regional and Systematic (or General) geography-the former focused on the study of particular places, and the latter was devoted to the study of the nature, and pattern of spatial distribution of particular items of geographical interest over the earth surface and its parts.

The most creditable part of Varenius's contribution lay in that he underlined the relationship between the two streams of geographical scholarship: Special geography provided the results of in-depth study of particular places and regions which became the raw material (the data) on the basis of which General geography could pursue its task of depicting spatial patterns of distribution, and inferring therefrom general hypotheses and laws explaining why they occur where they do, and thereby providing valuable inputs for better work in the area of Special (i.e., Regional) studies. Varenius pointed out that while Special geography was of great practical value in the pursuit of government and commerce, General geography provided information on the principles governing the distribution of particular phenomena on the earth surface so that the administrator and the businessman may be suitably informed about the nature of the environment they are likely to encounter in particular parts of earth's surface. To Bernard Varenius, therefore, General and Special geography did not suggest a dichotomy or a separation of ways, and division of objectives. To him, the two represented mutually interdependent parts of geography as a unified field of scientific learning. In this vision of geography, Varenius was far ahead of his peers. This explains why he had so greatly influenced the concept and scope of geography in Europe for well over a century.

In the foreword to his book, Varenius had set out a plan for Special geography, according to which the description of particular places should be based on celestial conditions, including climate; terrestrial conditions, including relief, vegetation and animal life; and human conditions including trade, settlements and forms of government in each country being studied. It is true though, that Varenius was none too enthusiastic about human geography since its subject matter could not be put to exact mathematical analysis for purposes of generating laws of behaviour (Gettfried Lange, 1961, paraphrased in Holt-Jensen, 1980, p. 14).

Like most other great works of scholarship, Varenius's book had been inspired by the demands of his time. In 1647, Varenius had accepted the position of a private tutor in a family in Amsterdam, then the commercial hub of the Netherlands. Here he came in contact with merchants engaged in international trade. Many of the merchants needed information about Japan where the Dutch had established a trading post in Nagasaki. This is what had inspired his first book entitled *Regional Description of Japan and Siam* published in 1649. The experience gained in writing a regional geography of Japan gave Varenius the idea that descriptions of particular places "could have no standing as contribution to science so long as these are not related to a coherent body of general concepts". His *Geographia Generalis* was written with a view to promoting the search for and the building of this much-needed conceptual coherence in geographical scholarship. His book went through several editions in Latin-two of these (published in 1672 and 1681) edited by no less a person than Sir Isaac Newton. An English edition was published in 1693 (Baker, 1955a and 1955b).

Varenius passed away in 1650 at the tender age of 28 so that the world of scholarship was deprived of many more conceptual leads. Underlining the methodology of Special vis-a-vis General geography, Varenius pointed out that while in General geography (dealing mostly with phenomena of physical origin), most things can be proved by mathematical laws, in the case of Special geography, with the exception of celestial features (i.e., climate), things must be proved by experience that is, by direct observation through the senses (James, 1972, p. 226).

THE IMPACT OF DISCOVERIES

New Answers to Questions about the Origin of the Earth and Its Surface Features, and Man's Place in Nature

Speculation about the origin of the earth, and man's place in the web of nature, had for long remained constrained by theocratic domination of thought in mediaeval Europe. Intellectual thinking had continued to be conditioned by traditions inherited from ancient Greece as well as from biblical accounts. All this began to change during the seventeenth century when steps were initiated to cut the thought process loose from the strangle hold of biblical beliefs, and to start experimenting with rational methods, so that geographical exploration had "immense significance in the history of science and of thought" (Parry, 1981, p. 3).

By the end of the seventeenth century, a good deal of speculation on the origin of the earth had led to the belief that the earth is a physical phenomenon that has acquired its present form though natural processes of change spread over millions of years, and that it was wrong to regard it as a divine creation. Inspired by the theory of comets given by Edmond Halley (1656-1742) in 1682, William Whiston (1667-1752) developed the theory that the earth was made from the debris of a comet, and that the gravitational pull of a second approaching comet had caused the elliptical orbit of the earth around the sun, and had led-through the tidal waves caused by its gravitational pull-to the creation of the continents and ocean basins. While the crests of the waves were occupied by landmasses (continents), the troughs became the ocean basins. Later the German scholar Abraham Gottlob Werner (1749-1817) developed the theory that the great flood that had been caused by the cooling of the earth's atmosphere, had led to dissolution of materials of the earth's crust. The dissolved materials of the crust were deposited on the surface of the earth in the form of a series of layers so that large parts of the earth surface are covered with sedimentary strata.

Simultaneously, a good deal of speculation had begun on the origin of landforms. In 1719, John Strachey (1671-1743) showed that landforms reflected the rock structure lying underneath them. Subsequently, in 1777 Simon Pallas (1741-1811) published geological maps to show that the cores of most mountain ranges are made of granite. Alongside, ideas about the mechanics of river flows and valley development were being developed at a rapid pace. The French scholar Louis Gabriel Comte Du Buat (1734-1809) mathematically explained (in 1786) how the flowing water of a river can establish equilibrium between velocity and the load of sediment being transported by it. This had ledto the idea of "graded river profiles". During the 18th century James Hutton (1726-1797) popularized the concept of uniformitarianism, according to which the processes that shape the earth surface indicate a perpetual process of change "with no vestige of a beginning and no prospect of an end".

New methods of scientific classification of plant and animal life were also influential in shaping geographic thought and practice. The most influential figure in this field was Swedish botanist Carolus Linnaeus (1707-1778). He developed a system of classification based on classes, orders, genera, and species. French naturalist Lamarck (1744-1829) drew attention to the need for a system of classification of plants and animals in accordance with their natural characteristics. He challenged the widely believed notion that plants and animals were created in their present form. Thus, he presented the rudiment of a theory of evolution that was later advanced and refined in a big way by Darwin and others who laboured to explain the mechanism through which the process of evolution of life forms had taken place. Coupled with the idea of uniformitarianism of Hutton, the theory of biological evolution had greatly impressed geographers about the role of time in the evolution of landforms**.**

The early part of the eighteenth century also witnessed the first beginnings of scientific study of man. The German scholar J.P. Sussmilch (1707-1767), in a book published in 1741, had demonstrated the existence of statistical regularities in population data. His research showed that the ratios between the sexes had remained nearly balanced, and that birth and death rates could be predicted on the basis of past trends. (However, the idea that numerical information about individuals tends to group around averages in accordance with the theory of probability) was put forward by Lambert Quetelet (1796-1874) only much later in 1848.)

As knowledge about the lands and peoples over the earth's surface increased, so did speculation about the role of the environment in shaping human behaviour patterns. French philosopher Jean Bodin (1529-1596) was one of the first to present a major work on this theme in 1566. Placing belief in the Greek concept of climatic zones, Bodin formulated the theory that people in the southern parts of the world (being under the influence of Saturn) are religious by nature; those living in the northern regions (living under the influence of Mars) were endowed with martial characteristics; and only people living in the middle regions (owing to the influence of Jupiter) were able to evolve a civilized way of life and live under the rule of law. English geographer Nathanael Carpenter in his *Cosmography* (1625) further advanced Bodin's idea regarding climatic zones and their influence in shaping human behaviour.

From these early beginnings of what may appear to us today as unscientific speculation about the man- nature relationships, progressively evolved more rational scientific analyses based on detailed observations and comparative case studies. In a piece published in 1719, the French scholar Abbe de Bos established a definite relationship between the weather and suicide rates in the cities of Paris and Rome. His analysis showed that in Paris, suicides were most common in the period before the onset of winter and just after the end of winter. In Rome on the other hand, most suicides had occurred in the two hottest months in summer (Glacken, 1967, pp. 556-558). Until the 19th century the most influential scholar who worked on this theme was Charles Louis Montesquieu (1689-1755). In line with the scientific ideas current in his time, Montesquieu wrote that warm climates favoured growth of despotism and slavery, whereas colder climates encouraged democracy and freedom so that, according to him, democracy tended to increase in direct proportion to increase in distance from the equator. Despite these crude observations on the relationship between man and the environment, Montesquieu was far from a crude determinist (Kriesel, 1968); he had given due allowance to human initiative and technology in reducing environmental constraints to human progress.

The progress in scientific thought through the newly acquired habit of questioning everything in sight represented a new tradition in scholarship. As James (1972) wrote, all these efforts were "new" in the sense that they offered new hypotheses, new methods of classification, and new ways of making use of mathematical principles of explanation. In the development of this new way of thinking, the ground breaking work was performed by the French scholar Count Buffon (George Louis Leclerc, Comte de Buffon, 1707-1788), who was director of the Jardin du Roi botanical garden in Paris from 1739 to 1788. By virtue of the position that he held, Buffon had access to a large collection of specimens of plants and animals, and to descriptions written by travellers and explorers. His forty-four volume work on *Histoire Naturelle, Generale et Particuliere* (1749-1804) (written in active collaboration with a large number of scholars) "represents one of the first works resulting from the reports of voyages of discovery in which attention was turned from oddities and marvels to a search for regularities and for the laws governing processes of change. His approach was nonmathematical and ... strictly inductive ... aimed at finding some kind of order in the flood of new infornation" obtained from the explorations and discoveries (James, 1972, p. 136).

While Buffon subscribed to the idea of a divinely created earth, he rejected the theory that the final plan of creation was in the mind of the creator and as such there was no need to look for causes of earth phenomena. Buffon was the first to focus attention on man as an agent of geographic change. He developed the idea that the earth has been cooling gradually, and that part of the warmth on the earth surface was derived from its hot interior. Buffon subscribed to the theory of climatic determinism inherited from the ancient Greeks but he was positive that man was not a passive agent, and that he was capable of adjusting to any climate through his technology and culture.

Inclusion of panels of trained scientists in the voyages of discovery, beginning around the last quarter of the seventeenth century, had greatly promoted scientific knowledge about the earth. The first such scientific traveller during 1698-1708 was the English astronomer Edmund Halley (1656-1742), the great scientific genius at deriving order out of complex data. He was the originator of the mortality tables in 1693, as also of many graphic methods for showing geographical distribution of physical features of the earth. His maps and discussions of the trade winds of the Atlantic (1686) provided the first illustration of wind directions and wind shifts. He also prepared the first map of magnetic variations using isogonic lines in 1701.

The father-son team of Johann (1729-1798) and George (1754-1794) Forster had accompanied Captain James Cook on his second voyage to the Indian and the Pacific oceans. In the course of this voyage, the two made botanical observations. It was in the course of this voyage that George Forster found out that the patterns of temperature on the eastern and western margins of landmasses are very different so that there was similarity between the climates of Western Europe and the western coast of North America. George Forster later played the pivotal role in attracting Alexander van Humboldt to geography. Another great scientific traveller of this period was Major James Rennel (1742-1830). He was one of the founders of the science of oceanography, and had served as the Surveyor General of India during 1767-1777. His *Atlas of Bengal* (1779) had gone through several editions, and it had remained a standard work of reference until around 1850.

The growing spirit of inquiry had, by the last quarter of the eighteenth century, egged on many scholars to seek scientific answers to the age old questions regarding man and his life upon the earth. A most prominent name in this regard was that of Thomas Robert Malthus (1760-1834) who published his famous essay on population in 1798 in which he set out his theory about the interdependent relationship between increase in population and food supply. He noted that population increases in geometrical progression whereas food supply grows only in arithmetical progression. As a result, population keeps on increasing until subsistence level is reached, so that its further increase is checked by famine and epidemics. At one place in his essay Malthus had used the phrase "struggle for survival", a term which, several decades later, was to inspire Charles Darwin (1809- 1882) in his explorations toward the theory of evolution of species through the process of natural selection. In his studies Malthus had showed that increase in agricultural production could not cope with the natural increase in population, irrespective of technological inputs. He was also the first to formulate the economic law of diminishing returns from increased employment of capital and labour (Dikshit 2018).

UNIT - 2: DIFFERENT SCHOOLS OF MODERN GEOGRAPHY

GEOGRAPHY AFTER HUMBOLDT AND RITTER: DEVELOPMENTS IN GERMANY

THE INTELLECTUAL CLIMATE OF THE TIME

The deaths of Humboldt and Ritter in 1859 marked the end of an era in the development of geography as a branch of scientific knowledge. Humboldt and Ritter had stood at the crossroads between the classical and the modem. In that sense they had jointly represented both a beginning and an end. The scheme of classification of knowledge presented by Kant (though not necessarily originated by him), had by now become commonly accepted so that geography had come to occupy a definite place among the sciences as a branch of knowledge which brings to bear a spatial perspective to the study of diverse phenomena on the earth's surface. The rapid explosion of knowledge following many a voyage of scientific discovery to far-off areas of the earth surface, and the analysis of the massive data collected there from, gave rise to a series of systematic sciences, each defined in terms of the kind of phenomena it studied. This represented Kant's logical division of knowledge, as contrasted to his scheme of physical division in terms of the arrangement of phenomena of diverse origin in time (history) or space (geography). The dual scheme of division of knowledge into a series of systematic sciences each devoted to a distinct subject matter, and its classification in terms of the association of diverse phenomena vertically in terms of time periods of history, or horizontally in terms of geographical space, implied a final fracturing of the classical view about the unity of knowledge. Humboldt was the last great figure who could claim universal scholarship.

The scientific voyages of discovery in the eighteenth and early nineteenth century had completely revolutionized the biological sciences. Diverse theories about the nature and origins of plants and animals in various parts of the earth's surface had begun to be pieced together to produce overarching theories, culminating in Charles Darwin's *The Origin of Species* whose publication in 1859 (coinciding with the death of Humboldt and Ritter) had contributed to a major change in perspective in every field of science-natural and physical, human as well as material.

Darwin's *The Origin of Species* established that organisms on the earth's surface have evolved through a slow and cumulative process of change; and that the evolutionary process of change among different organisms had not resulted from need or use (as postulated by Lamarck). Referring to Lamarck's famous example of the giraffe, Darwin noted that the giraffe did not get his long neck by stretching for fodder. The fact was that the individual giraffes that were born with longer necks were better equipped to survive in the Savanna environment than were the short-necked ones, and so this feature was slowly passed on to later generations. In his explanation of the mechanism of the evolution of species Darwin demonstrated that this was far from a straight line process, and randomness and chance had played a major role in evolution. Confirmation of the role of randomness and chance in the evolution of life forms cut at the very base of the long and widely subscribed view that the universe was a divine creation. Teleology could no longer be sustained as a scientific hypothesis, and was quickly abandoned. An equally important and radical concept was the idea of cumulative change through time. These ideas which appear as simple knowledge today, were highly revolutionary at the time they were presented, and were influential in changing the thought processes of contemporary science (Livingstone, 1992, pp. 178-189).

An aspect of the evolutionary theory, of particular importance to geography, was that it emphasized the need to study phenomena on the earth's surface in relation to the environment in which they were located, since only in this way could their struggle for survival and environmental adjustment be adequately appreciated. This gave a further justification for the methodology and perspective of geography as a science.

The concept of evolution, which attracted many scholars in branches of knowledge outside of biology, appeared in the study of landforms as Davis's cycle of erosion and was also reflected in the concept of mature soils developed through a slow process of change from parent materials. Applied to the study of human groups, it became the theory of social Darwinism of which Friedrich Ratzel was the staunchest supporter in geography.

THE CRISIS OF IDENTITY IN GEOGRAPHY

A major aspect of the changed perspective in science in the eighteenth century was the rise of systematic sciences. This meant that the subject matter that had previously been identified as general geography now became divided into the domains of a series of natural and physical sciences, as a result of which geography was faced with a real crisis of identity. Humboldt had tried to resolve this crisis by emphasizing that after the newly emerging systematic sciences had divided up the original content of general geography, there was still a field of study that had not been appropriated by any of them. This related to the study of interconnections among phenomena of diverse origins existing together in harmonious relationship in particular segments of the earth's surface. These interconnections give personality to particular areas and regions, and only through reference to such interconnections could the nature and spatial distribution of diverse earth phenomena e.g., plants and animals, and elements of climate etc. be properly understood and explained.

The most characteristic feature of the many systematic fields of sciences that had emerged in the eighteenth and early part of the nineteenth century was the method of study adopted by them. Their approach to study consisted of first isolating the processes or the phenomena that each of them examined, and then proceeding to formulate an ideal (abstract) model of how each process works in isolation. In other words, progress in the systematic natural and physical sciences was achieved by specifically excluding the effect of the real life situation wherein phenomena of diverse origin and therefore belonging to the domains of separate sciences, exist together in symbiotic relationship. Viewed this way, the methods and objectives of the systematic sciences appeared to be at cross-purposes with the holistic perspective of geography. Since geography could not identify any particular circle of facts as its special object of study, it began to lose its claim as a science irrespective of the unquestioned personal stature of Humboldt as the leading scientist of his time.

As the study of physical aspects of the earth was parcelled into the domains of a series of independent sciences, each seeking to isolate processes pertaining to the particular category that formed its special field of enquiry, the concept of "natural units" (Lander) comprising physical environment and man in particular segments of the earth surface as envisaged by Forster and further refined by Humboldt and Ritter, lost respectability as a field of scientific interest. Geography, therefore, began to be slowly neglected as a scientific pursuit, so that Ritter's chair in geography at Berlin remained vacant for a long time.

Geography experienced revival as an academic discipline only after the unification of Germany under the leadership of Prussia in 1871 following realization by the leaders of the government of the need to promote geographical knowledge about lands and people around the globe with a view to serving the interests of colonial expansion in the increasingly competitive world of imperialism.

Under the circumstances, geography was faced with three major tasks: Continued collection of information about the relatively little-known parts of the earth and its presentation in a useful form; detailed study of particular places in order to facilitate the work of officers of government, and serve the needs of businessmen and military commanders; and formulation of concepts about spatial association and variation of phenomena on the earth's surface through empirical generalizations.

Another significant and noteworthy aspect of the empirical-theoretical model of scientific research, promoted in the course of the rise of new systematic branches of science, was the insistence that the study of human and non-human phenomena required very different methods of study, and as such no field of scientific research could include both. In the intellectual climate of the times doing so was against the logic of science. The net result for geography was that after the deaths of Humboldt and Ritter in 1859, geography rapidly lost respectability as a branch of science. This led to a temporary parting of the ways between physical and human-regional geography; since the former could be pursued through the analytical method of natural and physical sciences, whereas the latter was descriptive and holistic in perspective. Serious attempts at reconciling this dualism between physical and human, and regional and systematic geography were made only in the 1880s.

DEVELOPMENTS IN GERMANY

For reasons indicated in the preceding paragraphs, after the death of Ritter, geography in Germany lost unified focus as a scientific discipline. Questions regarding its nature and methodology were seldom asked because they were difficult to answer. However, practical utility of geography to the army commanders in regard to the scene of military operations, to the government administrators in the newly acquired German colonial possessions, and to businessmen for purposes of trade and commerce was clearly recognized. For each of these purposes, clearly designed maps incorporating the latest available information were an urgent need. Accordingly, regional description and cartography became identified as the primary task of geography at that time. For the time being geography became "anything that could be put on maps". At the higher levels, as a field of teaching and research geography remained chiefly concerned with providing background information about the land and people in particular places with a view to facilitating a better comprehension of their history. All this began to change, however, after 1871 when the Prussian government (conscious of the utility of geography as an aid to colonial administration and territorial expansion) created a number of new professorships in geography in several universities across the country as a result of which geography regained respectability as a field of learning. A brief discussion of the ideas and works of some of the leaders of German geography in the last quarter of nineteenth century follows.

RISE OF DUALISM BETWEEN PHYSICAL AND HUMAN GEOGRAPHY

OSCAR PESCHEL (1826-1875)

For about twenty years before his death in 1875, Oscar Peschel had been the leading academic geographer in Germany. He raised some fundamental questions concerning the nature of geography, and was critical of the approaches of both Humboldt and Ritter. Peschel is described by some as the last great geographer before the discipline was finally overtaken by the impact of Darwinian ideas. Although most of Peschel's own works had appeared in print after the publication of Darwin's *The Origin of Species,* the implication of Darwinian ideas in the interpretation of earthbound phenomena and human societies had not yet been fully realized.

Peschel (1879) attempted to establish physical geography as a science. He was critical of Ritter for having neglected physical geography, and he criticized Humboldt for not having attempted a scientific classification of landforms. He was also critical of Humboldt for creating the impression that general geography could be equated with the entirety of natural science. Ritter was further decried for holding a teleological view about the nature and origins of earth phenomena, and for having subordinated geography to history by presenting his own works essentially to serve as a tool for better understanding of history. Peschel believed that physical and human geography constituted two entirely separate domains of knowledge, and as such the two could not form part of a single science. He was, thus, the originator of dualism between physical and human geography, and was strongly of the view that geography should be pursued primarily as a study of the physical phenomena of the earth. His methodological view regarding the distinction between physical and human in the study of geography notwithstanding, Peschel himself made an in-depth study of the geographical distribution of the races and cultures of mankind. It is indeed ironical that his book entitled *The Races of Mankind: Their Geographical Distribution* (1876) is the only work of Peschel available in English translation today (Dickinson, 1969, pp. 56-58).

GEORGE GERLAND (1833-1919)

Peschel's exhortation that in order to project its status as a science, geography should identify itself more and more as physical geography, was further advanced by another contemporary named George Gerland. Gerland was professor of geography at the university of Strasbourg, and had been the supervisor of the doctoral thesis of Alfred Hellner. In a long essay published in 1887 Gerland had gone so far as to suggest that since man could not be put to scientific analysis, study of man should be taken outside the purview of geography.

Thus, the continuance of geography as a unified field of academic inquiry incorporating the study of human as well as physical elements of the earth's surface was in serious crisis.

In the last two decades of the nineteenth century, several attempts were made to resolve this dualism, and to project geography as an integrated science concerned with the study of phenomena on the earth's surface both physical as well as human-in terms of their spatial distributions and areal associations. The most successful of such attempts were those put forward by Ratzel, and Alexander von Richthofen. Although Richthofen was the elder of the two, the first volume of Ratzel's methodologically epoch-making work *Anthropogeographie* (1882) was published a year earlier than Richthofen's epoch-making methodological statement presented in his 1883 inaugural address for the chair in geography at Leipzig. We discuss the two in that order.

RE- ESTABLISHMENT OF GEOGRAPHY AS AN INTEGRATED SCIENCE: THE STUDY OF MAN-LAND RELATIONSHIPS

FRIEDRICH RATZEL (1844-1904)

Friedrich Ratzel, described as the greatest single contributor to the development of geography of man.

The first fruit of his labour came in the form of the first volume of *Anthropogeographie* in 1882. The appearance of this book marked a major event in the history of geography in that it had a direct and far -reaching effect on the nature and methodology of geography which at that time was faced with a serious crisis of identity with regard to both its content and practice. According to Hartshorne (1939, p. 90) the term *anthropogeographie* was in itself misleading in that it suggested the connotation of geography of man in terms of individuals and races (that is, anthropological geography) whereas the major thrust of the work concerned the works of man, particularly the products of man's social life and relationship to the earth. (This was clearly a reflection of the influence of the ethnographer Moritz Wagner whom he had met in Munich.) With his sound background in the life sciences, Ratzel saw in geography the sought-for opportunity for establishing a connection between natural sciences and the study of man. Ratzel set out to lay a scientific foundation for the study of man in geography. In his *Anthropogeographie* he demonstrated that the geography of man and his work could as well be put to systematic analysis as the elements of the non-human (i.e., physical) world. Ratzel was, therefore, in the true sense of the term, the father of modern human geography as a field of scientific enquiry. The second volume of Ratzel's *Anthropogeographie* was published in 1891. The two volumes had represented two different approaches to the study of the human element in geography. The first volume was organized in terms of physical features of the earth, which were studied in terms of their influence on human culture. The central focus of this volume was to analyze how far and in what manner man's life upon the earth is shaped by the physical forces of nature. This was the common procedure adopted by geographers at that time, though some of Ratzel's contemporaries (most notably Kirchhoff, 1833--1907) had started studying human geography by the reverse method; that is, by analysing human activities and human cultures in relation to the physical environment which, in methodological terms, implied proceeding in the study of human geography by starting with man rather than the natural environment. The second volume of Ratzel's *Anthropogeographie* was written from this reverse perspective. Strangely, however, in the English-speaking world, it was Ratzel's first volume rather than the second that became the dominant input in human geographical methodology so that he came to be identified by later generations of geographers with the concept of human geography as the study of the works of man in terms of influences of the physical environment. This was the source of the long and erroneously held belief that Ratzel had advocated a deterministic view of human geography.

INTEGRATION THROUGH THE CONCEPT OF CHOROLOGY

Ferdinand von Richthofen (1833-1905)

As Hartshorne wrote, Richthofen's exposition of the relation of systematic (i.e., general) geography and regional geography with each other, and to the field of geography as a whole, was of immense importance in the development of geographic thought. The real purpose of systematic geography, according to Richthofen, was to lead to an understanding of the causal relations of phenomena in areas-an understanding which can be expressed in the forrn of principles that can be applied in the interpretation of individual regions, called *chorology* or regional geography. Richthofen distinguished between a first step chorology, which is non-explanatory description, providing material for systematic geography; and chorology, as a final step, in which the explanatory study of regions is based on systematic geography. Thus, as Hartshorne (1939, pp. 92-93) wrote: "Not bound by any limited concept of science, Richthofen, like Humboldt, saw no objection to a single science considering different kinds of things that exist together and are bound together by causal connections". To him geography involved no duality between physical and human; it studied both insofar as they illuminated an understanding of the earth surface as the home of man.

Alfred Hettner (1859-1941)

Geography, according to Hettner, begins with the spatial association of phenomena that give character to areas at different scales of resolution. Geography involves consideration of causal interdependence *(Zusammenhang)* of various sets of spatially arranged phenomena in particular segments of the earth's surface; individual distributions by themselves are of little significance in giving character to areas and as such are of little direct concern to geography as a specialized discipline. However, single distribution becomes important in characterizing an area insofar as it is coincident with other areal phenomena with similar areal distribution i.e., to the extent that the distribution of single elements is geographically (or spatially) efficacious. Like Richthofen, Heitner also used the chorological view of geography as a means to an end-to resolve the problem of dualism between physical vis-a-vis human geography. To quote him, geography as a chorological study "is neither natural nor human ... but both together", for it is concerned with the "distinctive character of lands", since the character of any region with a history of human settlement incorporated both physical as well as human aspects, both types of elements formed equally important components of geographical study. He criticized the likes of Peschel and Gerland who wanted to project geography as a purely physical science. He was also critical of Ritter and his followers who wanted to make man the central focus of geography. Hettner emphasized that man must be considered alongside nature in the study of areal descriptions and interrelations. Hettner addressed the question whether geography should confine its vision to the study of geographical areas as unique entities, or also concern itself with the formulation of general concepts and theories i.e., is geography idiographic or nomothetic in perspective? His own answer was, that like all other fields of scientific knowledge geography must be concerned both with the uniqueness of particular phenomena as well as the general principles that explain the particular examples. Geography is concerned both with the unique character of particular areas as well as the elements of similarity or universality between them. The in-depth study of particular regions can be effectively pursued only when it is suitably illuminated by the relevant general concepts which would frequently be necessary to identify their uniqueness. Likewise, the study of areas as unique assemblages of diverse phenomena provides raw material for general studies, since every unique element is a challenge that leads to the search for better theory. Richthofen's programme of geography, and its further exposition by Hettner paved the way for studies in regional geography interpreted in terms of the fruits of systematic geography. In methodological terms, this represented a return to Humboldt's approach to geography, an approach that was neglected by the followers of Ritter who pursued regional geography with relatively little concern for systematic geography. It is true that Hettner consistently encouraged the study of regional geography, but that should not be interpreted to imply that he discouraged systematic geography because he consistently emphasized the need for geographic work from both the perspectives. As Hartshorne (1939, pp. 93-95) wrote, it is significant to note that regional study in geography in Germany had hardly reached a position comparable with systematic studies until after the First World War which led to a refocusing of interest on the geographic character of different parts of Europe in general, and the Germanic lands in particular. The standard approach to regional study in geography, which had served as the general framework for many scholarly works in Germany published from the 1880s to the 1920s, was described by Hettner in an article in *Geographische Zeitschrift* (1933). Under this framework, various categories of facts were examined in their geographical distribution starting with geographical location, geology, relief, climate, natural resources, development and distribution of population and settlements, the forms of economy, trade, transport, and political divisions. This outline was based on the belief that the framework formed a kind of sequence of cause and effect. The framework had continued to be pursued as the model approach to regional study in geography in Germany for a long time. Through Hartshorne's *Nature of Geography* it became the standard framework for regional geography in the English-speaking world.

Although the chorological view of geography had become the dominant concept of geography in the working life of Hettner, it is significant to note that the chorological concept of the discipline had only *upstaged* and not *replaced* the Ratzelian view of geography as the study of man-land relationships. Indeed, in many parts of the English-speaking world, the Ratzelian concept had continued to be vigorously pursued under the influence of the highly pursuasive writings of Ellen Churchill Semple.

GEOGRAPHY AS A LANDSCAPE SCIENCE

Otto Schluter (1872-1952)

In a critical appraisal of Schluter's contribution, Leo Waibel (1933) wrote: Through the concept of the physiognomic build of the landscape , he gave the geography of man a corporate substance for research, which can be worked out according to the same method as physical geography-the cultural landscape. This can be examined from the standpoints of its morphology, physiology, and developmental history, just as the visible phenomena in nature in the build of the landscape. Between physical geography and the geography of man, there is no longer a gap. Both are in the closest contact in terms of objects and methods (cited in Dickinson, 1969, p. 132).

The concept of landscape and the related methodological principles of Schluter had been widely used by German geographers before the Second World War. Writing in 1952, German geographer Lautensauch reported that most geographers in Germany followed Schluter in identifying the study of landscape as the central task of geography: The German methodologists viewed every landscape as a “dynamic structure, a thing area- time system of specified quality inside the whole geosphere” that is, an open system, as contrasted to the closed system of organism" (cited in James, 1972, p. 232).

Jean Brunhes was advocating a somewhat similar set of ideas in France, but it had made very little impact in the English-speaking world until Carl Sauer introduced the concept in his essay on "The Morphology of Landscape" (1925), and later built a research group (the Berkeley School) that focused on the historical-ecological study of cultural landscapes. This concept was completely ignored in Britain so that Darby's (1936) edited volume made only a passing reference to Brunhes. Even in the United States the concept was not pursued outside the Berkeley School. Part of the reason lay in the widespread influence of Hartshorne's *Nature of Geography* (1939) which had convincingly argued in favour of the Hettnerian paradigm of geography as the time-honoured concept of the discipline.

Geography after Humboldt and Ritter:
Developments outside Germany

Geography had been firmly established as an academic discipline in Germany following large-scale endowment of chairs in geography in many universities across the country. The revival of geography, however, took some time to spread to other parts of Europe, and to the United States. There was, in general, a lapse of a decade or more. Before long, geography had begun to appear as a subject of higher learning in most European universities. The part played by the universities in the formation of geography as a professional discipline was of fundamental importance since only through the training of the younger generations in an accepted mode of study could geography emerge as a professional field.

DEVELOPMENTS IN FRANCE

The immediate cause for the revival of geography as a university level subject of study in France was the country's defeat in the Franco-Prussian war of 1870-1871. Soon after the war there was a great demand for better quality geography education in the French schools, with a view to stimulating interest in the knowledge about peoples and places in far-off lands so that the pursuit of colonialism could be attended to more effectively. The country's defeat, and the loss of territory, made it urgent that France should look to Africa and elsewhere for commercial opportunities and politico-cultural colonization (Freeman, 1971, p. 46). The development of French colonial power after 1871 owed a great deal to the influence of the French geographical societies (McKay, 1943).

Contributions of Vidal de la Blache (1845-1918)

Unlike in Germany (where revival of geography in the last quarter of the nineteenth century had been marked by the rise of several distinct schools of thought), the growth of modern geography in France was shaped by the work of one man, Paul Vidal de la Blache, who founded a new school of thought in human geography that remained dominant until the Second World War. Writing in 1922, W.L.G. Joerg had noted that "Nearly all occupants of chairs in geography in France are pupils or pupils of pupils of the late Vidal de la Blache. In no other country ... the development of geography centred about one man as in France".

Blache had come to geography via history and literature. After a doctorate in history in 1872, he began teaching geography, first at the University of Nancy (1872-1877), and later as professor of geography at Ecole Normale Superiere in Paris. He became the first geographer to be appointed to the chair of geography at the Sorbonne in 1898. As a contemporary of Ratzel, it was quite natural that Vidal de la Blache should be influenced by his writings. It was not Ratzel of the first volume of *Anthropogeographie* (1882) but Ratzel of the second volume (published in 1891) that attracted Vidal. He was drawn to Ratzel's concern with geographical distribution of man, and the role of migration (and inherited traits) in man's adjustment to nature. This basic concept became the central theme in Vidal's own concept of *possibilism,* which held that nature sets limits and offers possibilities for development, but the way man adjusts to the natural conditions of the area of his inhabitance is largely a function of his own tradition and mental structuring. The same environment carries different meanings to people with different *genres de vie* (ways of living or culture). According to Blache, culture (i.e., inherited traits) is the basic factor in determining which of the many possibilities in the natural environment shall be selected by a given community.

Blache was opposed to the concept of dichotomy between natural and cultural aspects of geography. To him, consideration of natural and cultural aspects of the earth's surface cannot be separated from one another for the simple reason that in every inhabited part of the earth's surface, the original landscape is significantly transformed as a result of human habitation. Such changes are greater in the case of culturally advanced societies where, owing to the more developed technology, the degree of man's intervention in nature is more far-reaching. It is, therefore, impossible to study landscapes meaningfully without due reference to the interlocking roles of nature and culture.

The relationship of a community to the physical landscape of the area of its inhabitance is so intimate that it is difficult to think of the one without the other. As such, each segment of the cultural landscape has, with the passage of time, acquired a unique personality of its own. Such areas were named as the *pays.* According to Blache, the study of such regions constituted the primary task of geography as a professional field. Elaborating upon his idea about the method of geographical enquiry, Blache (1913) stated that geography is "the study of things associated in areas, mutually interacting, characterizing particular segments of the earth space". According to him, the distinguishing feature of geography as a science was its "capacity not to break apart what nature has assembled, to understand the correspondence and correlation of things" existing together in association in particular regions or pays.

Blache's method of study was essentially inductive and historical. It was designed to suit the study of small areas of self-sufficient economies as had existed in the nineteenth century rural France. An agricultural way of life pursued in relative isolation from the outside world had favoured the development of locally distinctive traditions and ways of life *(genre de vie),* including agricultural practices, implements, food habits, dress and architecture. Blache recommended that geographers should carry out research in folk cultures with a view to depicting the unique personality of each region.

La Tradition Vidalienne

For almost half a century French geography had remained a faithful reflection of the Vidalian tradition of thought and methodology perpetuated by his own students and the next generation trained by those students as they spread out to man the various university departments across the country. The essence of the Vidalian tradition was an unswerving faith in the principle of terrestrial unity so that the concept of dualism between man and nature, and human and physical geography, was alien to the French tradition in geography. French geographers have all through maintained a balance between physical and human components of geography. Thus, of the two leading disciples of Blache (each a leader in his own right), while Jean Brunhes (1869-1930) spread the message and methodology of new geography of man as preached by his master, Emmanuel de Mortonne (1873-1955), was a leading physical geographer.

Brunhes was a leading geographer in his own right. His book *La Geographie Humaine,* first published in 1910, went through several editions over the next quarter of a century. An English translation (edited by Bowman) was published in 1920. The book became so popular in North America that an abridged edition was published in 1952. Brunhes proposed a classification of geographic facts that made Blache' s concepts easier to transmit in the classroom. Brunhes wrote that the two basic maps for the study of human geography are a map of water and a map of population. According to him the essential facts of human geography may be classified into three groups namely

• Facts of the unproductive occupation of the soil (houses, roads, and settlements);
• Facts of plants and animal conquest (cultivation of plants, and animal husbandry);
• Facts of destructive exploitation (clearing of forests, hunting, and mining).

The last part of Brunhes's book contained a discussion on different kinds of geographical studies pursued under the headings of human geography, regional geography, social geography, political geography and historical geography.

Emmanuel de Mortonne was a leading physical geographer of his time. He held physical geography to be an essential part of the scheme of geographical study of areas. He maintained a consistent interest in geomorphology and climatology. He combined this interest with regional expertise in the geography of Central Europe. One of the most influential geographers of the interwar period in Europe, de Mortonne was a strong supporter of Davisian geomorphology which he popularized in the French academic circles. His 1927 study on the identification of arid"regions through the use of aridity index, was a major contribution to the study of climate. An important part of the Vidalian tradition was the recognition that while from one point of view geography is a unitary field, from another it appears to tie together a variety of fields in the natural and human sciences.

In a book entitled *Science of Geography* (1925) Camille Vallaux (1870-1945) stated that geography is both a unitary and autonomous field of study, and also an auxiliary aspect of many other fields of scientific knowledge. Thus, not only does geography have a philosophy of its own, "it is almost, in itself, a philosophy of the world of man". This explains how the Vidalian tradition had succeeded in making important contributions to systematic aspects of geography (i.e., topical studies) side by side with its unique contribution to regional geography pursued through the concept of *genre de vie* which had resulted in the justly famous French regional monographs "the study of places" (Harrison-Church, 1951; James, 1972; Livingstone, 1992). Another distinctive aspect of the Vidalian tradition in geography was that Blache had regarded field study of small areas as the best possible way to train geographers. He was convinced of the great practical value of such regional studies to society and government. With a view to fulfilling this need, Vidal had planned a series of books on regional geography covering the whole of Europe a new series of universal geographies. However, Blache passed away before the plan could be concretized so that the task was taken over by his student and close associate Lucien Gallois (1857-1941). The first volume appeared in 1927, and the series had been completed (except for the volume on France) before the outbreak of the Second World War. The series is regarded as a monument to the regional tradition in geography.

DEVELOPMENTS IN GREAT BRITAIN

In the nineteenth century British geography had suffered from maladies similar to the ones that had beset it in France. Much of what was taught in British schools in the name of geography was uninteresting and dull; and the students were required to commit to memory a large mass of unorganized facts. As in France, courses in geography in the universities were handled by geologists on the physical side, and the courses meant to provide geographical background to history were taught by historians. The overall scene was extremely confusing. In the midst of this confusion Mary Somerville (1780-1872) appeared as a bright star and shone on the path to development of geography as a field of scientific enquiry in its own right.

Her famous book on *Physical Geography* was published in 1848. It went through seven editions over the next thirty years. This book started with a physical description of the earth's surface continents, oceans, atmosphere, plants and animals-and included the study of man as an agent of change in the physical landscape. The author kept on updating the book with each edition by incorporating new facts as they became available. Its sound methodology and approach received praise even from Humboldt (Freeman, 1971, p. 28) but irrespective of its so and method and the author's erudition, the book failed to create a stir in British geographical circles since it was published at a time when physical geography was claimed by geologists, and Humboldt's books were supposed to have said everything that needed to be told about geography. Ironically enough, the book became highly influential in North America through

George Perkin Marsh who found her observations about man's destructive use of the earth very stimulating. Another major figure in the history of nineteenth century British geography was Francis Galton (1822-1911) who was a member of the Council of the Royal Geographical Society in London from 1854-1893. He had prepared the first weather map of Britain in 1861, and was the first scholar to demonstrate that weather patterns could be revealed by plotting lines of equal air pressure on the map. He was also the first to identify the nature of air circulation in an anticyclone formation. He further refined his isopleths technique to prepare the first ever isochronous map showing lines of equal travel- time from London. Commenting upon the nature of geography Gallon (1855, p. 81) described it as "a peculiarly liberalizing pursuit, which links the scattered sciences together, and gives each of them a meaning and significance of which they are barren when tl1ey stand alone" (cited in James, 1972, p. 257).

While geography was making news through the contributions of individual scholars like Somerville and Gallon, until the 1880s the discipline had no place as a branch of learning in the British universities. In 1884 the Royal Geographical Society asked its secretary, John Scott Kellie to survey the status of geography in Great Britain in the light of the contemporary status of the discipline in Germany, France, and U.S.A. Kellie reported that the status of geography in Britain compared very unfavourably with its status in other leading European countries and in America where the subject was flourishing, as a university level discipline under the charge of full professors of geography. On the initiative of the society (and induced by a special grant offered by it), the university of Oxford introduced geography in 1887 with Halford J. Mackinder (1861-1947) as reader and chairman. The university of Cambridge followed suit in 1888 under Francis Henry Hill Guillemard (1852-1933) who was succeeded by John Young Buchanan in 1889.

As was the case in Germany and France, in Britain also expansion of geography as a university-level discipline was intimately related to the needs of British imperialism. A memorandum issued by the Royal Geographical Society had urged the promotion of geographical knowledge in Britain since The colonies of England, and her commerce, her emigrations, her wars, her missionaries, and her scientific explorers bring her into contact with all parts of the globe, and it is, therefore, a matter of imperial importance that no reasonable means should be neglected of training her youths in sound geographical knowledge (Proceedings of the Royal Geographical Society, 1879, pp. 261-264; cited in Freeman, 1980, p. 35).

Contribution of Halford J. Mackinder (1861-1947)

Through his writings and addresses, Mackinder became a major influence in British geography. In his presidential address to the Geography Section (E) of the British Association of Science in 1895 under the title "Modern Geography: German and English", he stated that while British geography could be proud of its contribution to field survey, hydrography, climatology, and biogeography, as regards the "synthetic, philosophical and ... educational side of the subject" the British contribution fell "so markedly below the foreign and especially the German standard". The main points in Mackinder's view of geography as a field of learning, as evidenced by his presidential address may be summarized as follows. On the relationship between systematic and regional studies in geography Mackinder stated that in his view "treatment by regions is a more thorough test of the logic of the geographical argument than is the treatment by types of phenomena". In this regard he commended the effort made by Alexander von Humboldt who "for the first time" made "an exhaustive attempt to [causally] relate ... relief, climate, vegetation, fauna and the various human activities" in particular places.

Mackinder did not in any way undervalue the contribution of the systematic approach to geographical inquiry. He noted that regional surveys could be made on the basis of systematic geography. Since scientific analysis of the environment depended on geomorphology, "geophysiology" (oceanography and climatology) and biogeography, anyone studying human geography in the regional context needed the knowledge of the environment gathered by systematic studies of its different aspects. Not aware of the concept of cultural borrowings through migrations in contemporary France, Mackinder drew attention to the role of migration as an instrument of geographical change, noting that although human communities were influenced by the geographical environment, their traditions and practices in relation to the use of the environment are derived from cumulative wisdom resulting from past experience so that by exercising ingenuity, human groups are able to maintain themselves even in the midst of very unfavourable natural environment: The student of geography must remember, therefore, that man's progress on earth is influenced by two sets of factors natural environment, and cultural tradition forming part of man's mental equipment. Mackinder emphasized the need for geographers to pay due attention to the economic aspects of life in particular places as an essential component of human geography. He also laid emphasis on the study of political geography. He regarded politics as an expression of the corporate spirit of a nation (here the influence of Ratzel's theory of the state is obvious). His political geography was based on economic and strategic considerations (see Freeman, 1980, pp. 51-53; and Freeman, 1971, pp. 65-66).

Contribution of Patrick Geddes (1854-1934)

Another geographer who made a lasting impact on British geography was Patrick Geddes. In an essay published in 1898 he expressed views about the nature of geography comparable to those of Mackinder (1895), but from a biological perspective, as contrasted to Mackinder's historical perspective. A botanist by training, Geddes was greatly influenced by the Darwinian idea of "ordered evolutionary unity". He had been attracted to geography with a view to investigating the possible relationships between relief, climate, and natural resources on the one hand and the distribution of various communities, their economic pursuit, and cultural development**,** on the other. He was deeply interested in finding an answer to the question: How has nature determined, and how has man reacted to his environment? For an answer, Geddes focused attention on the study of human activities (Robson, 1981). Influenced by the ideas of the French sociologist Frederic le Play (1806-1882), Geddes developed a methodology for the study of human communities through focused attention on the "Place-Work-Family (Folk)" progression: The essence of his methodology was that place (i.e., Environment) determines the pattern of economic life, which, in turn**,** determines family norms and social structures. In the case of certain communities of advanced cultures, the steps could be reversed: Folk-Work-Place. Geddes was a great advocate of learning and imparting education through regional surveys. His famous slogan was "Survey before action". Thus, "Geddes has remained a source of fascination to many people, and his ideas on town planning long survived his death in 1934" (Freeman, 1980, p. 53).

DEVELOPMENTS IN RUSSIA

The vast expanse of the Russian empire was a most potent factor in the development of geography as an institutionalized discipline. Peter the Great (who ruled Russia from 1682 to 1725) appreciated the need for accurate geographical information to facilitate the eastward march of the empire. State supported expeditions were sent to the east and the north to explore the vast uninhabited stretches of territory; and generous funding was provided to prepare maps of the explored regions. The main objective of these explorations was to chart out the topography, and to identify places where valuable merchandise, such as furs and precious metals could be found. Later on M.V. Limonosov (who became head of the world's first officially named Department of Geography at the Russian Academy of Sciences in 1758) pursuaded the state authorities to charge the exploring parties with the task of collecting systematic information about the physical character of land, population, and economic condition. The recognition of geography as a department in the Russian Academy of Sciences gave considerable academic prestige to geography as a useful field of scientific learning. Under the patronage of the Academy, the Department of Geography launched several schemes of regional surveys and mapping of data. In the beginning, many of the more important projects in exploration and mapping were done under the guidance of experienced experts from Germany and France, but gradually they were replaced by the newly trained Russian personnel. German geographer Busching was a pastor of a German Lutheran Church in St. Petersburg from 1761-1765. Portions of his *New Geography* dealing with Russia were translated into Russian, and his suggestion that the imperial territory be divided into natural regions in order to facilitate administration was quickly adopted so that by 1800, regional descriptions of a number of the regions had already been published. By the beginning of the nineteenth century Russian geography had already developed two distinguishing characteristics. The first was an emphasis on regions as the basis for organizing geographical work, and the belief that regions are concrete entities that can be objectively defined. Second was the continued-use of geography to include a wide variety of specialities. This was sharply contrasted to the contemporary trend in Germany where, in the words of James, classical geography was torn apart as each academic discipline established its separate existence. In Russia, the classical tradition of geography as a field of study dealing with the physical environment of the earth and its human inhabitants had continued. To formalize this structure of geography, the Imperial Russian Geographical Society was founded in 1845. The Society was charged with promoting work, alongside geography, in geology, meteorology, hydrology, anthropology, and archaeology. The several specialities covered under the Society were collectively identified as "the geographical sciences". Unlike in Germany where the deaths of Humboldt and Ritter in 1859 were marked by a break in the continuity of geographical study, in Russia there was no such break. For this reason, in the case of Russian geography, it is difficult to pick up any single scholar as the "grand old man of Russian geography". However, the continuous growth of the pre-Soviet (before 1917) period suffered a serious jolt, like everything else in the country, after the 1917 revolution. In order to survive the persecution of the new regime, scholars were required to bring their ideas in line with the ideas of Marx and his followers, who were strong advocates of economic determinism, and were clearly opposed to any suggestion that natural environment was a potent force in shaping the life of human communities in particular areas. Given this major curb on the exercise of free thought, in the post-1917 period, development of geography as a field of learning in Russia was cut off from the concept and development of the discipline in the rest of Europe and North America. This isolation became particularly marked owing to the strong linguistic barrier. Owing to this, developments in Soviet geography since 1917 have little bearing on the conceptual development of geography as a science. For this reason, the post-1917 developments in geography in Russia need not detain us. However, as regards the pre-1917 phase in the development of geography in Russia, Petr Kropotkin, whose concept of geography as social ecology, and his concept of "mutual aid" in nature have been the object of special attention in the recent literature (Galois, 1976; Breitbart, 1981), deserves special mention.

Contributions of Petr Kropotkin (1842-1921)

As Galois (1976, p .91) wrote, Kropotkin offers a view of nature and potentiality of geography which could replace the post-quantitative revolution "spatial monomania" with a socially conscious account of the states and conditions on the earth. That Kropotkin's contribution to the method and philosophy of social and economic geography had remained generally neglected until recently may be explained by the fact that Kropotkin had presented these ideas during a period when the capitalist nations of the West were attempting to consolidate their power over the resources of the world. Under the prevailing imperialistic accumulation of capital through colonial exploitation through the political strategy of divide and rule, Kropotkin's voice had no listeners so that when the Royal Geographical Society of London wanted to educate its members in geography through an arranged lecture series in 1887, it looked to young Mackinder who was then a young man of 26 shaping up into a brilliant speaker and scholar, rather than to Kropotkin who had by then already made a mark as an eminent geographer, and was at hand right in the office of the Royal Geographical Society.

DEVELOPMENTS IN THE UNITED STATES

Starting with the foundation of the Johns Hopkins University in 1875, the concept of university as a community of scholars had taken root in the New World. The time had now arrived for setting paradigms for scholarly performance in each discipline by its own professionals. The pioneering work in the introduction of the new geography, then making news in Germany and France, was performed by William Morris Davis who had taken appointment as an instructor in physical geography in the department of geology at Harvard in 1878. The first separate university department in geography to be established in the United States was at the University of Chicago in 1903 under the charge of the geologist Rollin D. Salisbury. Geography had, however, already been present in American schools and colleges for quite some time. Arnold Guyot (1807-1884), a Swiss scholar, and former student of Ritter, had arrived in the country in 1848 on invitation from the Harvard University to deliver a series of lectures outlining the nature of "new" geography. His lectures had been brought out in book form under the title *Earth and Man: Lectures on Comparative Physical Geography in its Relation to the History of Mankind,* the following year and had remained the standard reference on Ritter's ideas for a long time. Another early pioneer in this regard was George Perkins Marsh (1801-1882) whose book entitled *Physical Geography* *as Modified by Human Action* had been published way back in 1864. Through this book, Marsh had introduced the American public to the ideas of Humboldt, Ritter, and Mary Somerville focusing on interconnections between man and his natural surroundings. Another important name in this regard was that of Mathew Fontaine Maury (1806- 1873), a naval officer who had collected a mass of data on winds and currents of the oceans, and in 1850 had presented a model of atmospheric circulation on the earth's surface. His book entitled *Physical Geography of the Sea* (1855) was widely read.

Contribution of William Morris Davis (1850-1934)

Central to Davis's contribution to geography was the concept of "geographical cycle", as he chose to call it, though it is more popularly known as the cycle of erosion. His cycle was a generalized model of the sequence of landforms that would occur in the course of erosional work of running water on an elevated portion of the earth's surface, if there were no change in its surface in relation to elevation or sea level, and no drastic change in climate. His formula for landform evolution was based on a combined interaction between the *structure* of surface rocks, the agency of landform change *(process)* and the *stage* in the sequence of landform change already reached at a particular time. Davis belonged to the period in intellectual history when Darwin's ideas about the evolution of species were in great currency, and scholars in most fields of study were increasingly drawn to the concept of development through gradual and cumulative change over a long period of time (as part of the theory of evolution), and they were attempting to apply it as an underlying principle in their own researches. Davis's concept of cyclic development of landforms was clearly Darwinist in inspiration.

Contribution of Mark Jefferson (1863-1949)

As James wrote, Jefferson deserves a special place in the history of American geography not only because of the enthusiasm he enkindled in his students, but also for the many contributions to the conceptual structure of geography that came from his pen. Jefferson strongly disagreed with the recommendations of the Committee of Ten regarding the content and conceptual approach to geography. He insisted that the focus of geography teaching should be "man on the earth", in that order, not "the earth and man". Jefferson's geography (in sharp contrast to Davis's) was a man centred geography. Further, Jefferson was opposed to the view of Davis and the Committee of Ten which had favoured the study of systematic geography to the exclusion of regional geography of other parts of the world. As a professor in a teachers' training college, Jefferson realized the role of knowledge about lands and peoples around the world in training the mind of the future citizens of a rapidly shrinking world. Responding (in 1931) to a questionnaire in regard to the nature of geography, Jefferson stated that his view of geography was many things in one, so that it defied a short and crisp definition. To Jefferson, The nature of geography is the fact that there are discoverable causes of distributions and relations between distributions. We study geography when we seek to discover them. But there is an art of geography-the delineation of earth's features and inhabitants on maps-cartography, and a science of geography, which contemplates the facts, delineated and seeks out the causes of the form taken by each distribution and its relationship to others (cited in Martin, 1968, pp. 319-321; reference in James, 1972, p. 369).

Contribution of Elsworth Huntington (1876-1947)

Another former student of Davis at Harvard, Huntington was a most prolific writer and an imaginative interpreter of the effects of climate on human life. Correlating the periods of drought with historical dates, he developed the hypothesis that the great migrations and invasions of the nomadic peoples of Central Asia resulting in the Mongol conquests of China and India, and invasions over eastern Europe, in the thirteenth century, could be explained by dry climatic spells leading to the drying up of the pastures that formed the backbone of nomadic economy. He presented this theory in his book entitled *The Pulse of Asia* (1907). The book was a great success in the world of scholarship, and quickly launched its author on the road to academic fame. He became established as the most knowledgeable person on the influences of climate on human history. Another book published in 1915 under the title, *Civilization and Climate,* advanced the thesis that civilizations could develop only in the stimulating climates of the temperate regions, and that the monotonous heat of the tropics had destined the people of those areas to live in relative poverty. Huntington's books were immensely popular among historians, sociologists, and medical students (owing to his statements regarding the close relationship between health and climate), as well as geographers, but at the time that Huntington was highlighting climatic influences on civilization in a deterministic style, American geography was fast moving away from the philosophy of environmental determinism. Owing to this, Huntington's influence on American geography-its philosophy and methodology-was rather limited.

Contribution of Ellen Churchill Semple (1863-1932)

Another contemporary geographer whose name deserves special mention as a leading figure in the formative phase of American geography was Miss Ellen Churhill Semple who, after a Masters degree in history, had gone to Germany to study geography under Ratzel at the University of Leipzig in 1891-1892, and again in 1895. Ratzel's ideas made a lasting impression on her historically trained mind so that she was inspired to study American history in relation to its earth conditions. The result was the publication of her first book entitled *American History and its Geographical Conditions* in 1903. The book was an immediate success and confirmed her status as an eminent teacher of history and geography. Semple's place in the historiography of American geography lies in that she had carried Ratzel's philosophy and methodology of geography to the New World, and had mesmerized a generation of American students by her persuasive writing and her enchanting literary style, backed by convincing scholarship. Through her generations of American geographers came to subscribe to the view of environmental influences in shaping man's life upon the earth. Her version of the first volume of Ratzel's *Anthropogeographie* was presented in her book entitled *Influences of Geographical Environment* (1911). The opening paragraph of the book set its tone and style:

Geography as Human Ecology: The Contribution of Harlan H. Barrows

In American geography, the concept of geography as human ecology was set forth in clear terms for the first time by Harlan Barrows (1877-1960) in his 1922 presidential address to the Association of American Geographers (published in the *Annals,* A.A.G. in 1923). Like Jefferson and others before him, Barrows underlined that man's adaptation to the conditions of his habitat is not caused by the physical environment. On the contrary, this adaptation was a function of human choice the central idea behind the concept of possibilism, then *so* much in currency in France. Barrows had expressed the view that although much of the subject matter of geography had been lost to the systematic disciplines, the content of geographical study was still far too broad for its development as a coherent discipline.

He, therefore, proposed that the specialized branches of geography, such as geomorphology, climatology, and biogeography should be abandoned, and geographers should concentrate upon human ecology as the unifying theme the central organizing principle for their work. Barrows was of the view that "those relationships between man and the earth which result from his effort to get a living [are] in general the most direct and intimate", and that most other human adjustments on the earth's surface are guided by these economic relations. Accordingly, he recommended that "development of economic regional geography should be promoted assiduously, and that upon economic geography for the most part, other divisions of the subject must be based". He emphasized the need for intensive field work and regarded "a thoroughly effective technique of field work as our greatest immediate need", since, in his view; the field represented the geographers' laboratory.

Geography as Chorology: The Contribution of Carl Sauer (1889-1975)

The concept of geography as a chorological science had a long history of development in Germany and the other countries of Europe. In America, however, geography had remained confined to the search for geographic influences in human development. By the beginning of the 1920s, many among the younger generation of geographers had started questioning the validity of the content and method of geography as a discipline built around the basic theme of man-environment relationships. The new generation of students was anxiously looking for an alternative model for geographical work to replace the existing one which they were finding increasingly unsatisfying. It was at this juncture in the history of American geography that Carl Sauer (1889-1975), who had then taken charge of a new department of geography at the University of California at Berkeley, published his frequently cited essay entitled "The Morphology of Landscape" (1925).

Sauer was critical of the concept of geography built around the theme of man-land relationships, and underlined the intellectual strait-jacketing that adherence to this concept as the guiding principle of geographical work had created, since the concept tied the student to a single dogma that committed him to a particular outcome of investigation in advance. The guiding premise was that the environment influenced man, and that in course of man's life upon the earth the environment itself experienced changes.

Sauer proposed an alternative model for geography based on the works of Humboldt and Hettner. Following the teachings of the two eminent German geographers, Sauer defined geography as the study of things associated in area on the earth's surface, and the differences in the nature of areal aggregations from place to place in regard to both physical as well as cultural factors. In particular, he drew attention to the role played by human action in modifying the physical and biotic features in particular segments of the earth's surface so that natural landscapes had got slowly transformed into cultural landscapes (Hooson, 1981).

Development of Historical Geography

Chorology was centrally focused on the study of present-day landscapes, and following the Kantian scheme of division of knowledge on a physical basis between chronological (i.e., historical) and spatial sciences-a view to which both Hettner and Hartshorne subscribed, chorology remained indifferent to the historical perspective which, according to the current interpretation of the Kantian view, belonged to the domain of another group of sciences. However, as students started applying themselves to the study of actual landscapes, they came increasingly to realize that the "being" and the "becoming" of landscapes were intimately interrelated; so that the present-day patterns in the landscapes of particular places could not be adequately appreciated without due attention to the processes (i.e., the time sequence of events) that had contributed to bring them to their present form. This meant that in practice, the chorologist was often required to go backwards in time to adopt a historical perspective. As a result of this realization, historical geography was soon raised to the status of an essential component of the American approach to the study of geography as chorology. The influence of Sauer in this regard was quite obvious. Needless to say, however, that the historical geography of Sauer and his contemporaries was very different in nature and orientation from the one that Ellen Semple had laid down in her enchanting *American History and its Geographic Conditions* (1903). The two chief innovators in historical geography of the post-1925 period in the United States were Ralph H. Brown, and Carl 0. Sauer. In his path-breaking study entitled *Mirror for Americans* (1943), Brown portrayed the geography of the eastern seaboard of the North American continent for the period around 1810 on the basis of statements contained in the writings of a large variety of contemporary authors. Brown's imaginative approach to re-creation of the past geography of a region, (in terms of how contemporary scholars had perceived it) had, in some ways, foreshadowed the modern orientation to environmental perception in geographical studies. In a second book entitled *Historical Geography of the United States* (1948), Brown had traced geographical changes that had taken place in the course of the country's settlement by European emigrants. The other major influence in the development of historical geography was Carl Sauer himself. He, and a large number of his graduate students at Berkeley, produced a number of studies of past landscapes. A major principle that emerged from these studies, focused upon sequences of the settlement process in particular areas, was that "the same physical conditions of land could have quite different meanings for people with different attitudes toward their environment, different objectives in making use of it, and different levels of technological skills" (James, 1972, p. 407). For example, in agricultural areas slope had one meaning for the man with a hoe and quite another for the man with a tractor-drawn plough. Likewise, people with one kind of culture might concentrate their settlements on flattish uplands, whereas another group belonging to a different cultural tradition might concentrate its dwellings in the valleys. Derwent Whittlesey was another major contributor to historical studies in geography. In a note published in the *Annals,* of the Association of American Geographers in 1927, he described studies focused upon the processes of change in the settlement of an area as studies of *sequent occupance.* According to Whittlesey, each generation of human occupance is linked to its forebear and to its offspring, and each exhibits individuality expressive of mutations of some elements of its natural and cultural characteristics. Moreover, the life history of each discloses the inevitability of the transformation from stage to stage. Clearly enough, sequent occupance represented the antithesis of the concept of environmental determinism. In the opinion of Preston James, the sequent occupance studies represented a kind of *cultural determinism* in that they had started with the premise that with every significant change in attitudes, objectives, and technical skills of the inhabitants of an area, the resource base of the region is put to fresh appraisal.

Development of Applied Geography

As large numbers of trained geographers poured out from the new university departments, it was natural that their young enquiring minds should question the practical utility of their knowledge in society's efforts to improve the quality of life. On this score, they found their discipline to be severely wanting. They noted that field studies in small areas carried out by students as part of the requirement for a formal degree were of little practical utility. They expressed themselves in favour of reorienting geographical education in a manner that would make it more relevant to the resolution of overriding social, economic and political problems of the day. In response to this rising demand, a large number of studies of an applied nature carried out with a view to providing valuable inputs in planning and remedial action were taken up and their results published. Thus, geography took on an applied perspective. As a result geographers now began to find employment as consultant S-in both public as well as private sectors. The breakthrough in the use of geography professionals in the study of practical problems had come as a result of the First World War in the course of which geographers had rubbed shoulders with other specialists to find solutions to real-life problems. In the decades following the First World War, all through the 1920s and 1930s different kinds of applied research were undertaken. In the Second World War, a large number of geographers had taken up appointments in the army and civilian departments, and there was growing appreciation of their contribution. The trend had continued (Dikshit 2018).

UNIT - 3: CONCEPT OF PARADIGM SHIFTS IN GEOGRAPHY

Although 'the appearance of a "new geography" has been loudly and persistently proclaimed by almost every generation of geographers since ancient times', or so Preston James (1972,505) tells **us,** the last decade has witnessed the novelty of geographers seeking support for their claims from studies in the history and philosophy of science. Thomas Kuhn's *The Structure* of *Scientific Revolutions* has received special mention. Despite the frequent references to Kuhn's work, however, fundamental questions about the implications of his ideas for geography remain unanswered. One of these questions is prompted by the repeated appeals to Kuhn's scheme as support for programmes of reform in the discipline; a second arises from the lopsided treatment of his work in which his concept of 'normal science', the necessary corollary to scientific revolution, has been totally neglected; a third comes from the lack of any serious attempt to analyse the history of geography in terms of his scheme.

Has the dominantly prescriptive use of the paradigm concept in geographical writing been consistent with Kuhn's use of the term? What are the implications for geography of Kuhn's view of 'normal science'? Has the kind of scientific revolution envisaged by Kuhn really occurred in geography?

THE PARADIGM PANACEA

At one point in *The Structure* of *Scientific Revolutions* (p. 160) Kuhn mused that Probably questions like the following are really being asked: Why does my field fail to move ahead in the way that, say, physics does? What changes in technique or method or ideology would enable it to do so? Geographers certainly were and still are asking this question, and have seen in Kuhn's work a justification for their programmes for reform. Kuhn's ideas have apparently been welcomed *not* because of their value as analytical tools which might sharpen our knowledge of the development of geography, but rather because they have allowed recent trends in geography to be likened *to* phases of rapid advance in other, more scientifically respectable disciplines.

Rapid progress can be **(or** has been) made, we have been repeatedly told, by the adoption of some new paradigm or exemplar, By implication much traditional geographic work has seemed pre-scientific. Nonetheless, Haggett and Chorley (1967) have assured us, 'Geography, coming late to the paradigm race, has the compensating advantage that it can study at leisure the "take-off'' paradigms of other sciences'. What geography has been transformed by (or still needs) is a new paradigm such as model building (Haggett and Chorley, 1967), or systems analysis (Chorley, 1971, 1973; Chorley and Kennedy, 1971), or a 'massive shift from descriptive geography towards more analytic work in which mathematical models of how regions grow and interact play a dominant role' (Haggett, 1970, 20), or even a 'process meta geography that leads to a paradigm of locational and environmental decision making in complex systems' (Berry, 1973). Nor have the exhortations been limited to academic circles: the secondary school teacher of geography has been told both that he 'must become aware of the paradigms he has internalised as a result of his educational experiences', and that his teaching will be greatly improved by 'selecting a (geographical) paradigm to achieve educational aims' (Biddle, 1976).

My concern here is not with an evaluation of these programmes for reform - they all may be laudable - but with the impression that they give of deriving support from Kuhn's study of the development of science. I will argue that the prescriptive use of his concepts in geographical writings has been unjustified (see also Guelke, 1971). To use a concept as an analytical tool is one thir.g; to proclaim it as a panacea is quite another. In Kuhn's own words

. . . the proto-sciences, like the arts and philosophy, lack some element which, in the mature sciences, permits the more obvious forms of progress. It is not, however, anything that a methodological prescription can provide . . . I claim no therapy to assist the transformation of a proto-science to a science, nor do I suppose that anything of the sort is to be had. If , . . some social scientists take from me the view that they can improve the status of their field by ***first legislating*** *on* ***fundamentals and then turning*** *to* ***puzzle solving, they are badly misconstruing my point.*** . . **As** in individual development, so in the scientific group, maturity comes most surely to those who know how to wait. (Kuhn, 1970b, 244-45, italics added; see also Kuhn, 1962a, especially 160-61; 1962b, 190)

Indeed the suggestion that a disciplinary matrix can be changed by such legislation is the very antithesis of Kuhn's argument. In any discipline, Kuhn argued, there are initially no clear cut rules of research, and fact gathering is a fairly haphazard activity usually restricted to data that lie readily to hand; though the practitioners are scientists, the result of their labour is 'something less than science' (Kuhn, 1962a, 13**).** Research, he continued, is transformed by an achievement sufficiently unprecedented to attract adherents from competing modes of work, yet sufficiently open-ended to leave a wide range of problems to be solved (see also Masterman, 1970). In his view of science shared examples serve the cognitive functions generally attributed to shared rules, and he asserts that such knowledge 'develops differently from the way it does when governed by rules' (Kuhn, 1974,482). The role of exemplars is not just a matter of practice in applying rules, it is rather a matter of acquiring the ability to see resemblances, of perceiving similarities that are 'both logically and psychologically prior' to any rules (Kuhn, 1974, 472).

According to Kuhn, scientific progress is not gradual, orderly, or necessarily cumulative, but rather **is** characterised by truly revolutionary changes as one research tradition or programme replaces another. Nonetheless, he specifically warns against the trap of seeing a paradigm as 'a quasi-mystical entity or property which, like charisma, transforms those infected by it' (Kuhn, 1974, 460-61). Yet this very trap seems to have ensnared some seeking new directions for geography.

NORMAL GEOGRAPHY?

If indeed there have been revolutions in geography there must also, according to Kuhn, have been periods of 'normal' science; the dominant role of shared examples ensures this. In fact the novelty and significance of Kuhn's work lie far less in his treatment of revolutionary; phases than they do in his comments on the periods between major advances. There has, however, been an almost total neglect of the concept of 'normal' science in geographical writing. Apart from assurances that all will be well, we have been told virtually nothing of what will happen after a Kuhnian revolution in geography. Normal science, as Kuhn, sees it, has three main aspects: it includes the pursuit of facts that the discipline matrix shows to be worthwhile, the comparison of factual data with predictions, and the articulation of the disciplinary matrix by the resolving of ambiguities. Such activities entail the solving of puzzles, and not the testing of the essential assumptions of a disciplinary matrix. Kuhn is adamant here, insisting that while the striving to bring theory and fact into closer agreement may be seen as testing or as searching for confirmation, its object is the solution **of** a problem 'for whose very existence the validity of the paradigm (read disciplinary matrix) must be assumed' (Kuhn, 1962a, 80). And he is by no means simply repeating the old contention that selection, evaluation and criticism are guided by theory; he goes much further, claiming that **'no** process yet disclosed by the historical study of scientific development at all resembles the methodological stereotype of falsification by direct comparison with nature' (Kuhn, 1962a, 77).

Challenges to a disciplinary matrix, Kuhn (1962a, 78-9) argues, may go unnoticed, while the discovery of anomaly leads not to an immediate rejection of prevailing ideas but rather to their defence by the formulation of ***ad hoc*** hypotheses. In fact, he continues, a new paradigm or change in the matrix may initially offer no solution to questions successfully answered by the prevailing one. One must 'have faith that a new paradigm will succeed with many large problems that confront it, knowing only that the older paradigm has failed with a few. A decision of that kind can only be made on faith' (Kuhn, 1962a. 158). Faith, he emphasises, is necessary because of the 'incommensurability' of competing views. The proponents of each are always at least slightly at cross-purposes. Neither side will grant all the non empirical assumptions that the other needs in order to make its case . . . the competition between paradigms is not the sort of battle that can be resolved by proofs (Kuhn, 1962a, 148).

Briefly then this is the nature of Kuhnian normal science and it is a far cry from the delightfully rational rendition of paradigm-directed research and cost benefit weighting of competing paradigms (Haggett, 1977) presented in geographical writings. Even if we accept it as a true picture of our discipline, our sailing may be none too smooth because claims that the rejection / acceptance of paradigms is generally not subject to critical assessment and is always based on more than a comparison of theory to fact, or that research problems are made meaningful only by paradigms whose validity must be assumed, have prompted many rejoinders. Shapere (1964) comments that '. . . anything that allows science to accomplish anything can be part of (or somehow involved in) a paradigm', while Suppe (1974) suggests that the concept has become 'bloated to the point **of** being a philosophical analogue to phlogiston'. Scheffler (1967) argues that in such a scheme 'independent and public controls are no more, communication has failed, the common universe of things is a delusion, reality itself is made by the scientist rather than discovered by him'. Berkson (1974) claims that belief in the necessity of suspension **of** criticism after the adoption of a paradigm can cripple fruitful critical activity, and also points out those significant contributions have been made by those who have no faith in the ruling paradigm. Popper (1968) maintains that the hallmark of scientific statements is that they are testable, not that they are grouped around a central paradigm. 'It is just a dogma (he insists) . . . a dangerous dogma . . . that the different frameworks are like mutually untranslatable languages' (Popper, 1970). Debate of these issues falls beyond the scope of my paper, but, and this is what the advocates of paradigm shift in geography have not made clear, it is precisely such criticism that geographers must face if they embrace Kuhn's theory (see Lakatos and Musgrave, 1970).

A KUHNIAN GEOGRAPHY?

The need for an assessment of the development of geography in the light of Kuhn's arguments is clear enough. In fact Preston James has already attempted such an assessment, but his handling of Kuhn's terminology in a history of geography which emphasises continuity and gradual transformation (James, 1972, especially pp. 227, 395, 505) is strikingly incompatibIe with the stress placed on discontinuity and revolutionary change in The *Structure* of *Scientific Revolutions***.** He tells us, for instance, that 'it was Alfred Hettner who elaborated Richthofen's concept of chorology into a paradigm for geographical study', then argues that 'when Hettner presented this view, it had apparently been accepted by so many generations ***of*** geographers that he felt no need to support it with references to earlier writings' (James, 1972, 226-27). Moreover the differences between the German school and *la tradition vidalienne* were nothing like the intellectual gulf implied by James's use of Kuhn's terms. If correct, and notwithstanding his use of that terminology, James's emphasis on continuity in the discipline demonstrates the irrelevance to geography of Kuhn's version of scientific revolution and paradigms. However James was content with listing notable achievements or changes of fashion in geography and made no real attempt to come to grips with the complexity in even the early version of Kuhn's scheme.

While James was apparently too hasty in cutting the geographical coat to the Kuhnian cloth, Harvey (1973) was no less so in using a Marxist cudgel to knock the stuffing out of a Kuhnian straw man. Harvey argues that the overriding weakness of this 'idealist's' interpretation is its neglect of the manipulation of nature in the interest of man (especially middle-class man), yet he ignores Kuhn's counter-argument that it is precisely the overwhelming concern with paradigm-directed research that may lead the scientist to neglect pressing social problems.

Clearly, caution will be needed in assessing the relevance of Kuhn's work to geography – just how much caution can be readily seen from one example which at first sight seems a likely contender for the title of paradigm (i.e. in the sense of a disciplinary matrix). W. M. Davis's work has had an undeniably profound influence, and has recently been portrayed as little less than the hub of the geomorphic universe (Chorley ***et al.,*** 1973). Nonetheless it seems on several counts to fall short of a revolution *a la* Kuhn. Firstly, it was essentially cumulative, involving no distinct break with previous work. The great watershed surely lies earlier, at the triumph of fluvialism over its rivals. Secondly, the Davisian approach was never universally accepted, nor were its opponents regarded as heretics beyond the pale **of** geomorphic decency. Thirdly, the recent waning of its importance was by no means the outcome of a fatal crisis; this much is admitted by one of its most trenchant critics (Hack, 1975). Instead we have witnessed an increasingly precise recognition of the spatial and temporal Smits **of** cyclicaI interpretations (for example, Schurnm and Lichty, 1965; Jennings, 1973, 1978). In fact the **growing** awareness of those limitations is apparent in the later works of Davis himself. Though the rise and subsequent decline of the Davisian school thus seem to have little in common with Kuhn's revolutions, they do provide excellent examples of the major attributes of 'normal science'. Davis's work undoubtedly did open up previously unthought of problems, ranging from intercontinental comparisons of erosion surfaces down to the enumeration of 'cycles' of river terracing. Most of these problems did take the form of 'puules' in which the validity of the cyclical model was simply assumed; the taunt of 'how many peneplains can sit on the top of a mountain?' (Tuttle, 1975), was not far wide of the mark. There were few attempts to thoroughly test the cyclical model. while the challenges that did arise were generally ignored or met by recourse to *ad* hocdefence: consider the fate of Crickmay's challenge; or, for some Australian examples, see Young (1978).

If caution **is** needed in the interpretation of Davisian geomorphology, how much more is required when we turn to the so-called revolutions of the last two decades or so. Admittedly much of the work done in this period has the same balance of faith and critical temper as Kuhn says is typical of 'normal science', while an enormous effort has undoubtedly gone into 'puzzle-solving'. Yet this same period saw the publication of Glacken's Truces *on* rhe *Rhodian* Shore, a work untouched by the 'new geography' but with few if any peers. The changes of the 1960s may well have been revolutionary, but they seem to have been less profound and widespread than was thought a few years back (for example, see Jeans, 1978, for comments on historical geography). Moreover, I hazard the guess that the new movements of the 1970s have more in common with the anti-scientific groundswell expressed in the writing of Roszak (1973) than they do with the scheme that Kuhn derived from the physical sciences.

Acceptance of Kuhn's argument that a disciplinary matrix develops essentially from the study of exemplars rather than from some explicit catechism must also throw doubt on claims that geography as a whole has been or is being transformed by one revolution or another. Where are the new examples shared by human and physical geographers, alike? Even the textbooks which presented the novice with an image of a unified geography seem to be going out of style. Changes over the last two decades have in fact cut us off from the shared examples we once had. For, with apologies to both Lewis Carrol and Dury (1970), is it merely from nervousness not from goodwill that we march along shoulder to shoulder?

CONCLUSION

Almost invariably the versions of Kuhn's work presented in geographical writings have been distorted. The prescriptive use of the 'paradigm' concept advanced in support of vaned programmes for reform in our discipline is at odds with his theory of the growth of science. This, of course, in no way detracts from the worth of these programmes, but it does require that they are stripped of any mystique that surrounded the term 'paradigm'. **A** thorough historical analysis must replace the mere listing of notable achievements before we will know whether geography has actually undergone revolutions in Kuhn's sense of the term. But I suspect that the periods of 'normal science', rather than the sequence of 'new' geographies, will yield the greatest returns **for** a Kuhnian analysis. For like Popper (1970, *52),* I can see the dangers of paradigm-directed research (Young, 1978), but also express an 'indebtedness to Kuhn for pointing out the distinction (from revolutionary science), and thus for opening my eyes to a host of problems which previously I had not seen quite clearly'

UNIT 4: DUALISM AND DICHOTOMIES IN GEOGRAPHY: PHYSICAL GEOGRAPHY AND HUMAN GEOGRAPHY, REGIONAL GEOGRAPHY AND SYSTEMATIC GEOGRAPHY, IDEOGRAPHIC APPROACH AND NOMOTHETIC APPROACH

MEANING OF DUALISM AND DICHOTOMY IN GEOGRAPHY

The word *'dualism'* simply connotes the state of being divided. For any domain of knowledge therefore, it means two conceptually contrasted stances. Dualism finally leads to *'dichotomy'* which means the bifurcation of any subject into branches of knowledge.

Ever since its inception as a domain of knowledge, geography has been encountered with several methodological issues that eventually gave birth to several dualisms and dichotomies in the subject. Such sort of dualism was prevalent even in the classical or medieval periods of geographical history. Greek scholars like Aristotle, Herodotus or Hecataeus emphasized on physical geography; Roman scholars like Strabo insisted on regional geography while Ptolemy stressed on mathematical geography; and, the Arab scholars like Al-Masudi, Al-Biruni or Al- Idrisi highlighted on the importance of the physical environment. However, such dualisms were very equivocal and abstruse.

It was in the post-Renaissance period that geography witnessed the evident rise of dualism and since then, the subject has been branched off into several exclusive domains on methodological grounds. Over time the divisions have been further sub-divided into different sub-disciplines.

Some of the most conspicuous dualisms known to have existed in geography were:

- General (Systematic) Geography versus Regional Geography.
- Physical Geography versus Human Geography.
- Historical Geography versus Contemporary Geography.

SYSTEMATIC GEOGRAPHY VERSUS REGIONAL GEOGRAPHY

The dichotomy between systematic and regional geography was essentially rooted in another dualism that existed in the *approaches to study geography.* This dualism was between the ***Idiographic or Inductive Approach*** and the ***Nomothetic or Deductive Approach.*** The dichotomy between the two approaches may be explicated as—the idiographic or empirical approach did not seek to develop laws but mainly focused on the description of particular places in the context of their lands, seas or places and attempt to find its relation with other places. The nomothetic or deductive approach on the other hand, sought to establish laws and made general deductions based on those laws.

Dualism in geography was formally introduced in the 17th century which is often described as the *classical period of modern geography* by the German geographer, **Bernhard Varenius**. Using the terms of **Bartholomew Keckermann** a German philosopher, Varenius in his ***'Geographia Generalis'*** partitioned geography into-

- ***Special geography*** essentially concerned with the description of particular places on the basis of direct observations. This branch of geography was assumed to have great practical importance for governance and commerce.
- ***General geography*** based on universally applicable mathematical or astronomical laws.

Gradually, general geography evolved into systematic geography by incorporating the methods of the systematics sciences, while special geography evolved into regional geography. In simple words, the two may be expounded as the study of the natural vegetation of the world is a systematic approach while the study of a continent with respect to its natural vegetation, landforms, climate etc. is a regional approach.

The prominent German geographer **Alexander von Humboldt** followed Varenius and laid the foundation of systematic geography. In his famous book ***'Cosmos'*** Humboldt asserted that geography was meant to understand the *'harmonious unity of the cosmos.'* He distinguished between ***uranography*** as descriptive astronomy dealing with the *celestial bodies* and, ***geography*** as dealing with the *terrestrial part* with the prime objective of deciphering the unity that exists in the vast diversity of phenomena. It was not only the natural phenomena that Humboldt spoke of but, he also asserted that there existed unity of the human races as well since all the races had a common origin and therefore, no race was superior to the other. The unity of the phenomena, a viewpoint that Humboldt obtained from the German philosopher **Hegel** was based on the conjecture that there existed coherence as well as some sort of causality among them. The understanding of that unity was supposed to be derived from an understanding of the unity that subsisted between humans and the physical landscape. In fact, Humboldt opined that like other phenomena, *humans were basically a part of the nature*. Knowledge of the natural or physical phenomena was categorized by Humboldt as:

- ***Systematic Sciences***: This included sciences like botany, zoology or geology that classified phenomena according to their form and grouped them on the basis of certain commonalities.

- ***Historical Sciences***: This dealt with the development of phenomena over time.

- ***Geography or Earth Sciences***: This concerned itself with the spatial distribution and spatial relationship and interdependence of phenomena. It included all earth phenomena whether organic or inorganic.

Humboldt in his Cosmos stressed on his views that, for a comprehensive knowledge of the cosmos it was necessary to pursue systematic studies of particular phenomena and their interrelationship with other phenomena rather than undertaking complete studies of specific areas.

According to **Carl Ritter**, a contemporary of Humboldt, geography was concerned with *'lokalverhaltnisse'* or local conditions which described a spatial unit on the basis of *three* characteristics---

- ***topographical***, concerned with the delineation of natural divisions on the earth's surface;

- ***formal***, which dealt with the distribution and movement of such phenomena as water, air etc. that constituted the bases of human life;

- ***material***, which dealt with the distribution of biotic life, minerals etc.

Ritter provided the above purpose of geography in his famous ***'Erdkunde.'*** It was Ritter who introduced the inductive method in geography. He sought to develop a regional geography for which he used *'erdteile'* or continents as his units of study. He was of the idea that all continents had similar physical features and thus divided each continent into a highland core drained by major rivers of the land and, a low-lying coastland at the periphery.

Thereafter, in the late 19th century, geographers were highly influenced by the Darwinian doctrine and made significant contributions in furthering systematic geography. The most prominent among them were **Ferdinand von Richtofen** and **Friedrich Ratzel**.

Richtofen perceived geography in the same line as Humboldt as, the science of the earth's surface as well as the phenomena on it that were causally interrelated with it. According to him, the purpose of systematic geography was to provide an understanding of the interrelationship and causality of phenomena on the earth's surface which could be used for deducing about individual regions as well. He provided a guideline for the systematic study of the earth's surface. Richtofen also differentiated between general or systematic geography as analytic and regressive that was based on general concepts and, special or regional geography as synthetic and descriptive dealing with the unique and peculiar.

Friedrich Ratzel in his ***'Anthropogeographie'*** set a framework for the systematic study of human geography and thus set a new trend in the subject. Prior to him, systematic geography only involved physical geography and, human geography was mainly confined within regional studies. His anthropogeographie was essentially a reflection of the Darwinian viewpoints and emphasized on the concept of natural selection that was used in the natural sciences. Ratzel was of the view that cultural differences of a land were much more prominent than the physical differences. Ratzel's concept of geography was based on *two* propositions---(i) the interrelation of environment and humans and (ii) the interrelations of humans.

Alfred Hettner distinguished between systematic geography as that which was interested in formulating general laws and theories and, regional geography as concerned with the study of peculiarities in which the generalisations were tested to improvise on the existing theories.

The regional tradition was again revived by the French geographer **Vidal de la Blache**. He introduced the concept of *'pays'* or small local units as ideal units of study for the geographers which could even be used to arrive at general conclusions. He was contested however, by **Reclus** with his ***'Le Terra'*** that was centered on systematic physical geography.

The dichotomy between systematic and regional geography subsequently led to the **Hartshorne-Schaefer** debate. While Hartshorne in his ***'Nature of Geography'*** advocated that geography was regional in its essence and put forward his concept of ***areal differentiation'***, his views were refuted by Schaefer as *'Hartshorian Orthodoxy'* who called for a systematic scientific approach for geographical studies.

PHYSICAL GEOGRAPHY VERSUS HUMAN GEOGRAPHY

It was **Varenius** again as one of the first scholars to highlight on the differences in the nature and content of physical and human geography. He himself however was not much interested in the latter since human geography could not be subjected to mathematical laws to generate universal principles. He believed that the methods of the natural sciences could be successfully used to draw conclusions about natural phenomena with precision to a considerable extent. But they could not be applied to human groups because they were more subject to probability than certainty. Following him, was **Immanuel Kant** who offered a regular course of lectures on physical geography between 1756 to 1796 at the University of Konigsberg According to Kant, physical geography not only included the features visible on the earth's surface created by natural processes but also by human actions. Kant opined that physical geography was the first part of knowledge of the world and was essential to develop the basic understanding of the earth as the abode of humans and for furthering philosophical studies.

After Kant, it was **Humboldt** who stressed upon the study of physical geography. Since he believed in the *'unity of nature,'* in his opinion, physical geography was the study of phenomena arranged on the earth's surface and mutually related to each other that constituted the *'natural whole.'* Humboldt was of the view that differences in the economic, social and political conditions of different spatial units were largely a function of differences in natural conditions. Thus, according to him, human factors were subordinate to the natural factors. On the contrary, by upholding the teleological view that sought to provide a philosophical interpretation for geographical phenomena, **Ritter's** view of geography was anthropocentric in nature. Ritter conceived the earth as created by God with a *'purpose'* to educate humans and facilitate in their development. Ritter's anthropocentric geography stated that the way natural phenomena of any spatial unit affected its inhabitants, the inhabitants could also have an influence on the land.

Under the impact of Darwinian ideas, geographers focused more and more on physical geography. In 1848, **Mary Somerville** authored her ***'Physical Geography'*** and in 1877, **Thomas Henry Huxley** authored ***'Physiography'*** as the study of nature. In the second half of the 19^{th} century, more and more geographers were inclined towards physical geography. It is believed that **Hettner** accorded greater importance to the physical environment in comparison to cultural environment. German geographer, **Albrecht Penck** coined the term *'geomorphology'* as the study of landforms and established it as a distinctive sub-field of physical geography. It was Penck who formulated the principles of *'landform evolution.'* He also highlighted on the importance of relief maps in the systematic study of geographical elements. Subsequently, American geographer **William Morris Davis** also put forward his *'normal cycle of erosion.'* There were other scholars like **Koppen, Martonne, Mill,** and **Dokuchaiev** who put greater emphasis on landforms or climate as the major focus of geography. **Semple** went forward to explain humans as *'product of the earth's surface.'* **Mackinder, Chisholm, Herberton** and **Huntington----**all of them recognized physical geography as the core of the discipline of geography. Over the years, several sub-fields of physical geography have evolved like *geomorphology, climatology, oceanography, pedology, biogeography and environmental geography.*

The human element in geography was formally introduced in the work of **Ratzel** which actually gave rise to the dichotomy between physical and human geography. Ratzel in his ***'Anthropogeographie'*** described geography as the study of humans in the context of different races. But, Ratzel too was influenced by the Darwinian views and incorporated two major Darwinian tenets in his works-----(i) *struggle and natural selection* and (ii) *association and organization.* He used these themes in drawing an analogy between the political units and living organisms and thus came to be known as the ***father of political geography***. However, it was the French geographer, **Vidal de la Blache** who may be regarded as the founding father of modern human geography. However, Blache was of the opinion that human geography was a natural science. He adopted an inductive and historical approach in putting forth his propositions of human geography. His ***'Principles de Geographie Humaine'***, had several parts each devoted to several aspects of human geography. The introductory part analysed the principle of terrestrial unity as well as the concept of cultural milieu. Part first focused on population clusters and density. The second part was a description of man-milieu relationship. The third part dealt with transport and communication which was completed later by Martonne who also added the components of human races, diffusion of innovation and cultural regions in it. The Vidalienne tradition was carried forward by his ardent follower **Jean Brunhes** who propagated Blache's views of human geography not only within France but in other parts of the world as well. French historian **Lucien Febvre** was also inspired by the Vidalienne human geography. He put forward that humankind emerged as a powerful agent of modifying the earth's surface through centuries of their accumulated labour and decision-making. American geographer **Isaiah Bowman** also championed the cause of human approach in geography. In 1924, American geographer **Carl O. Sauer** propounded his ***'landscape paradigm'*** in which he highlighted on humans as agent of *'fashioning'* the natural landscape.

Over the years, studies in human geography has led to various sub-branches in this field as------ population geography, settlement geography, economic geography, social geography, cultural geography, political geography, historical geography and so on.

The dichotomy that existed between two prominent philosophies of geography namely, ***environmental determinism*** and ***possibilism*** could be attributed to the dualism between physical and human geography. In fact, the two revolutions that geography underwent namely, ***the positive-quantitative revolution*** and the ***critical revolution*** were somehow related to this dichotomy. This may be justified by the fact that while the former attempted to introduce the methodologies of the systematic sciences in geography as had been mostly done in the field of physical geography, the latter developed as a critique of the former mainly emphasized on the *'humane'* essence of the subject.

HISTORICAL GEOGRAPHY VERSUS CONTEMPORARY GEOGRAPHY

That all history should be treated geographically and all geography historically was asserted by the Greek scholar **Herodotus.** Thereafter, **Immanuel Kant** opined that since any individual's experience was restricted to a specific time and space, his knowledge had to be supplemented with the experiences of others. Such knowledge derived indirectly from others could be divided into *two* types—(i) narrative or, (ii) descriptive. While history was narrative, geography was descriptive. Thus, history and geography made up the entire gamut of empirical knowledge—the former that of time and the latter that of space.

Therefore, the importance of historical approach to geographical studies was acknowledged. The dichotomy between historical and contemporary geography also came to be regarded as an important dualism in geography. Historical geography dealt with the geographical description of any spatial unit as it had been in the past. The work of **S. M. Ali,** ***'The Geography of the Puranas'*** in which he endeavored to provide a geographical account of ancient India, could be considered as a remarkable work in this field. The contributions of the American geographer, **Ralph Brown** in the field of historical geography was also of great prominence. In fact, the **Vidalienne** tradition also adopted a historical approach.

However, historical geography could not be treated as a conventional field of geography as of contemporary geography. Hence, limited works could be found in this domain of study. Nevertheless, it incorporated in it both systematic and regional studies and included all the possible aspects of geographical knowledge.

The scope and content of historical geography were centred on the following themes:

- ***The Geographical Factor in History***: In the latter part of the 19^{th} century historical geography was conceived to be associated with the study of the mutual relations of phenomena over space in a particular period of time or, to study the impact of geographical phenomena in shaping the history of a region. **Whittlesey** emphasized on the study of historical factors in geography as it was thought to provide a spatio-temporal framework to study any spatial unit.

- ***Changing Cultural Landscape***: Historical geography was also considered to be the study of the cultural landscape as it existed in the past in any area such as, the settlement or the cropping patterns, house types etc.

- ***Reconstruction of Past Geographies***: This was an important aspect of historical geography. Since it embraced all fields of geographical knowledge, reconstruction of past geographies was essential for contemporary geographers as it enabled them to interpret the geographical phenomena of any spatial unit in present times in a more comprehensive manner.

- ***Geographical Changes through Time***: The concept of space had always been a central focus in geography. Geographical phenomena over space, whether natural or cultural changed with time which in turn, changed the character of space. The study of these geographical changes with time was of utmost importance to the geographers.

Contemporary geography on the other hand included all the geographical knowledge of modern times. In fact, contemporary geography had also witnessed the emergence of the modern and post-modern era with a constant revamping of approaches and methodologies in the geographical discipline.

DUALISMS IN GEOGRAPHY---A MYTH OR A REALITY

It is a fact that methodological differences had given rise to several dualisms in geography but the question that arises is that, whether the dichotomies that resulted from such dualisms had produced exclusive fields of knowledge or whether they are mutually related and transcended into one another. To be the precise, the question is whether dualisms in geography were a myth or a reality.

Varenius who actually introduced the tradition of dualism in geography, while categorizing between general and special geography asserted that they were mutually interdependent branches of geography. He stressed on the fact that special geography provided the database based on which general geography could infer the general hypotheses and laws. Humboldt recognized the interdependence of areal phenomena and opined that to understand a whole comprised of multiple phenomena it was essential to have knowledge about the constituents of that whole. Similarly, though Ritter adopted an inductive approach he acknowledged the contributions of Humboldt's systematic studies that enabled him to undertake special studies of regions. Richtofen on one hand while attempted to follow the precedence of Humboldt in establishing the affinity between geography and the natural sciences, on the other hand he also tried to restore the Ritterian tradition. It was Hettner who actually removed the dualism of the idiographic and nomothetic approach in geography. He stated that geography could involve both idiographic and nomothetic methods.

Both systematic and regional geographies could be regarded as the two extreme points of a continuum that gradually merged into the other. In regional studies, the concept of *'compage'* was introduced by **Derwent Whittlesey** to explain that regional geography was not a mere description of phenomena characteristic of any spatial unit but studied the functional association that existed between human beings and their physical, biotic and social environment. Therefore, through these arguments the dualism that is known to have existed between systematic and regional geography seemed to have been blurred. The general could be deciphered only through the particular which in turn, was not independent of the general.

The dualism to follow was between physical and human geography. In this too, the basic question posed was whether humans could be studied in exclusion from nature. At the same time, natural landscapes were occupied by humans. It was not possible to study human phenomena independent of the natural landscape and natural phenomena without their relation with humans. So the major thrust was on the relationship between humans and environment that constituted the central thesis of geography. This relationship however underwent several modifications sometimes according greater importance to nature and sometimes to humans or placed humans in harmony with nature. In physical geography, while explaining the normal cycle of erosion or landform evolution analogy was drawn with the lifecycle of humans, the concept of *'pays'* in human geography involved small *'natural'* regions. Thus, physical and human geography instead of being in contrast rather complemented each other which in turn faded away the dualism between physical and human geography.

Finally regarding the dualism between historical and contemporary geography, it may be stated that contemporary geography would become a part of historical geography over time. To comprehend the present it was highly essential to know the past. Therefore, historical geography provided a base for studying contemporary phenomena and how they have evolved over time. Hence, **Mackinder** asserted that historical geography was basically the study of the historical present. So, even in case of the dualism between historical and contemporary geography it was evident that one eventually led to the other and hence, their dualism also stood as illogical.

UNIT 5: POSITIVISM AND QUANTITATIVE REVOLUTION IN GEOGRAPHY; HARTSHORNE-SCHAEFER DEBATE, SYSTEM APPROACH IN GEOGRAPHY

THE CONCEPT OF POSITIVISM

The origin of positivism as a well-established philosophy can be accredited to French philosopher **August Comte** in the 1830s. Positivism as a philosophy was mainly initiated as a polemical instrument against the romantic and speculative tradition that prevailed prior to the French Revolution. Its main purpose was to distinguish science from metaphysics and religion. Thus, positivism may be precisely described as a philosophical movement that emphasized on science and scientific method as the only source of knowledge and, which stood in sharp contrast to religion and metaphysics.

Comte rejected metaphysics for *two* reasons (i) its abstract nature with no grounding in

reality; and, (ii) for being more concerned with emotional than with practical questions. He sought for *'sociocracy'* dominated by scientists for the unity and progress of the entire humanity. Since a lot of social disorder was created following the French Revolution, Comte attempted to establish positivist philosophy as an organizational tool that would lead the society through an organized development. This was much against the metaphysical principle that sought to change society through utopian solutions to the existing situations. Therefore, Comte argued that philosophy was an *'immature science'* and metaphysics should hence be replaced by a scientifically dominated *'positive'* outlook.

August Comte delivered lectures on positivist philosophy which was published as a book with the title ***'The Course of Positive Philosophy.'*** His positivist philosophy included the following *five* basic guidelines:

- All scientific knowledge was to be based on **direct and empirically verifiable experience of reality (*phenomenology)***. This was supposed to provide an edge over theoretical conjectures. Scientific methods, he asserted, was to combine both reason and experience—reason to formulate the hypothesis; and, experience to do away with falsifications.

- There was to be a **unified scientific method or *'le certitude'*** acceptable to all the sciences. This implied that the different branches of knowledge were to be distinguished by their subject matter or the object of study rather than their method of study. In other words, branches of knowledge differed from each other not on the basis of how they studied but on the basis of what they studied.

- This was possible only when there was ***'le precis'*** or a common objective of **formulating scientific theories** that could be subjected to empirical testing and utilized for proposing universal laws. This meant that ethical-based value judgements (beliefs, customs, norms etc.) were not to be considered as part of scientific knowledge since they were not based on direct observations and thus, were not capable of empirical verifications.

- The empirically verifiable theories so developed were supposed to be based on the tenet of ***'le utile'*** meaning that it should have had some **utility** to serve as an **instrument of social engineering.**

- Finally, the positivist philosophy was supposed to follow the doctrine of ***'le relative'*** which implied that **scientific knowledge was never complete** but rather relative. It kept on progressing with time through the **unification of scientific theories** which increased human awareness about the social arrangements that in turn, required more inclusive theories.

The philosophy of positivism challenged several taboos and religious beliefs that existed against empirical investigations. The above five postulates provided some sort of transition from the immediate through the unitary to the universal.

Comte opined that development of the society took place in *three* stages:

- ***theological*** when everything was described as God's will;
- ***metaphysical***; and,
- ***positive*** when attempts were made to find out some sort of causal relations between the observed phenomena.

Comte advocated that it was true that the social phenomena were more complex than the natural phenomena, yet, there should have been a science of social relationships to be developed as parallel to and in the same principles as the natural sciences. The purpose of such social sciences would be to explore the laws governing human society through the scientific investigation of social communities. These ideas of Comte were much in tune with the proposition of **John Locke** that knowledge could only be derived through direct observations of actual situations and whatever were not supported by empirical facts could not considered as knowledge.

In a nutshell the basic tenets of positivism were:

- Positivism was also described as ***empiricism*** (derived from the Greek word *'empeire'* meaning experience) since it promoted science and scientific methods as a source of knowledge. It averred that science only dealt with *'empirical questions'* that were based on experiences of real conditions as they existed and that which could be tested through experiments or some other measures. It enabled to discover the causal connections between the facts to arrive into some conclusions that were supposed to be value-free, unbiased and unprejudiced.

- The positivist philosophy proved to be ***anti-idealistic*** that is, it stood in contrast to anything that was abstract and essentially a mental construct. Therefore, positivist philosophy did not deal with the *'normative questions'* since they could not be tested empirically and could not be established with scientific evidence. Positivism thus rejected metaphysics for being unscientific.

- Since positivism declared anything as unscientific and abstract until it could be verified with empirical evidence and tested through experiments, it did not accept authority just because it was declared as authority. This brought them in conflict with the Nazi Movement in Germany and positivist philosophy was branded as ***anti-authoritarian*** and the term *'positivism'* was used as an abusive term.

In the 1920s, positivism witnessed some sort of deviation from the classical Comtean ideas when, a group of scientists created the ***'Vienna Circle'*** and identified themselves as the ***'logical positivists.'*** German philosopher and physicist **Moritz Schlick** was the founder of this group which also had another German philosopher **Rudolf Carnap** as a prominent member. They upheld the viewpoint that some knowledge could also be gained without relying on experience, through formal logic and pure mathematics. Hence they distinguished between:

- ***analytical statements*** which were, a priori propositions whose truth could be verified through tautologies and, which were essentially the domain of the formal sciences like logic and mathematics; and,

- ***synthetic statements*** that were supposed to be established empirically through hypothesis testing and these in turn, were supported by the analytical statements.

Hence, logical positivism offered a much more authentic basis for scientific investigation. The essence of logical positivism was acquiring knowledge through a combination of both experience and analysis and using such knowledge to alter phenomena so as to yield a desired outcome. This philosophy included *three* interrelated precepts:

- ***Scientism***: This meant that the positive methods alone were the methods of acquiring knowledge.

- ***Scientific Policies***: This implied that only positivism was the key to social engineering or modification as it provided rational solutions to all social problems.

- ***Value-Freedom***: This implied that scientific judgements derived through positive methods were neutral, unbiased and objective and hence were free from any moral or political binders.

POSITIVISM IN GEOGRAPHY

There was a great deal of efforts in the latter half of the 19th century to develop the discipline of geography as a ***nomothetic science***. This was largely the impact of the **Darwinian tradition** that invigorated the scientists to search for the governing laws of nature and in the same tune, the social scientists to explore the laws determining social arrangements. The ***hypothetic-deductive approach*** of study that was especially characteristic of the natural sciences, replaced the inductive methods in the social sciences. Thus there was an effort to accommodate social sciences within the framework of positivism. It must be pointed out here that the geographical developments that took place in the 1950s and 1960s were mainly committed to logical positivism. The researchers sought to develop a priori models about reality for which they devised a set of hypotheses that were to be authenticated, validated or discarded through testing of empirical data. Once verified, they were validated as laws until their eventual refutation through further research. The logical positivists conceived that some order persisted in the objective world that needed to be explored and discovered through scientific investigation- spatial patterns of variation in geography----that could not be manipulated by the observer. Geography soon became *'positivist-led.'* The hypothetic-deductive approach led the discipline particularly human geography to develop as a model building and theoretical science since it dealt with phenomena that were familiar with reality both spatially and temporally.

During the 1950s and 1960s, the essence and purpose of Anglo-American geography witnessed a drastic transformation with the replacement of the ***idiographic approach that focused on areal differentiation*** with the adoption of the ***nomothetic methods that sought to explore models of spatial structure.*** This change was initiated by **Schaefer** with his critique of Kant's exceptionalist views that placed history and geography as exceptional and different from the other systematic sciences. Schaefer put forward his ***'spatial organization paradigm'*** and conceived geography as a spatial and social science primarily concerned with the formulation of laws that governed the spatial distribution of any phenomenon as they were found on the earth's surface. Hence, Schaefer set off a sort of ***'revolution'*** in geography that was basically ***'theoretical and quantitative'*** in nature. This revolution in geography sought to provide the discipline a scientific approach with the application of mathematical and statistical methodologies. It largely accepted the tenets of positivism of unified scientific methods acceptable to all the sciences----natural and human. The quantitative schools undertook to construct models and theoretical structures within which geographical realities were supposed to be incorporated.

The quantitative revolution that geography underwent by adopting the viewpoints of positivism was set off mainly by the mathematicians. It was mainly the outcome of the impact of the non-geographers on geography. As was evident in many other disciplines, it altered an already existing knowledge base with a mathematical approach. **William Bunge** in his ***'Theoretical Geography'*** **(1962)** went as far as to describe geography as a science of spatial relations and geometry as the mathematics of space. So logically, geometry was supposed to be the language of geography.

What followed was the development of the concept of ***'space'*** as the basic concept for organizing the subject matter of geography. *Two* major approaches were identified in the study of geography, namely----

- ***Spatial Analysis***: This referred to the application of quantitative or more specifically statistical methods and techniques in locational analysis.

- ***Spatial Science***: This concept was largely akin to the positivist philosophy and presented human geography as a social science with its prime focus on space as the guiding principle behind the organization and operation of the society and the behaviour of individual human being.

These two approaches related to space led to the development of the following *two* aspects of space that became the central theses of geography:

- ***Spatial Interaction***: This referred to the interdependence between spatial units and was gratuitous to the interaction between humans and environment within a particular area.

- ***Spatial Structure***: This referred to the spatial arrangements or more precisely the geometric pattern of any phenomena on the earth's surface since geometry was regarded as an important tool in geographical studies.

Inspired by positivist thinking, major advances towards a unified methodological and philosophical basis of the quantitative schools were rendered by **Peter Haggett**, **Richard Chorley** and **David Harvey** in the 1960s. The discipline of geography witnessed major theoretical and methodological developments. A new domain of knowledge emerged that came to be known as ***regional science***. It was basically an assemblage of geography, economics and planning with its main concern for regional problems. The pioneer for this new discipline was Walter Isard (1956).

The most important theoretical development that fundamentally incorporated the philosophy of positivism was the ***locational analysis*** of **Peter Haggett.** The concept was put forward by him in the book ***'Locational Analysis in Human Geography'*** (1965). Following the geometric tradition this approach in human geography, more popularly termed as ***spatial science*** concerned itself with, the spatial arrangements of phenomena on the earth's surface. In addition to this, it also dealt with the interaction between places within a spatial pattern, the dynamics of such patterns as well as the creation of alternative patterns through model building to provide for a better possibility. The **Central Place Theory** postulated by **August Losch** (1954), the **Gravity Model** by **Stewart** and **Warntz** (1959) or the **Diffusion Theory** of **Hagerstrand** (1953) were all formulated using locational analysis.

Another concept that was intrinsically associated with positivism was the **concept of systems**. A system was defined as an array of entities that had specific relationship among them as well as with their environment. **Richard J. Chorley** was the first geographer to introduce general systems theory in geography. His paper ***'Geomorphology and General Systems Theory'*** (1962) was developed within the framework of the systems approach in which he tried to apply the concept of open and closed systems to geomorphology.

A major contribution to the positivist theory was made by **David Harvey** in his ***'Explanation in Geography'*** (1969). He opined that reality was a set of complex phenomena particularly so far as the relationship between the phenomena were concerned. However, it was possible to decipher such complexities with the aid of **system analysis** which explored the structure and function of a system. Every system was supposed to have three fundamental aspects structure,

function and development. The structure of the system was the set of elements it was comprised of and the relationship between them; function was the exchange or the flow between the elements while development meant the changes in the structure and function of the system over time.

Since geography studied the relationship between humans and the environment, systems analysis was supposed to have a wide range of applications especially in human geography. This was because the systems analysis was based on an implicit assumption of positivist philosophy and drew analogies between human societies on one hand and natural phenomena on the other.

This drawing of analogies led to the **model and analogue theory** that had close connections with positivism. A model was basically a structured conceptualization of reality that represented particular attributes of reality and, analogue theory was the formal theory related to building of models. A model or an analogue ranged from a structured idea to a hypothesis to a law to a theory. Following the positivist outlook, a model could be used as a guide to validate a set of hypothesis through empirical testing and to establish a theory as it contained some resemblances with the reality. Though model building had been used in many sciences since long back but, its use in geography was of comparatively recent origin and could be attributed to positivism.

CRITIQUE OF POSITIVISM

The positivist philosophy by rejecting metaphysics provided a sound philosophical, methodological and scientific base to the discipline of geography. Knowledge based on the observations of real situations that could be easily verified empirically was highly objective, unbiased and unprejudiced and could be readily utilized for the formulation of universal laws and theories. Positivism encouraged the use of statistical and mathematical techniques that provided precision to research and enabled to analyse a geographical system in a much more simplified form. It provided a kind of framework within which theoretical statements could be formally presented. However, the critique of positivism was highly intense and convincing. Positivism was criticized mainly on *three* fronts:

- ***Empiricism***: Positivism recognized the fact that theory building was essentially based on the direct observation of reality which could be subjected to statistical procedures for empirical verification. But this approach proved to be very superficial against which new philosophical and theoretical frameworks were designed and for which alternate methods than statistical techniques were required. These, like realism and structuralism offered a much more insightful exposure of human society. In contrast to positivism that concerned itself with *'how'*, they were concerned with *'why'* and went beyond the positivist argument to discover the processes that created a particular pattern of physical or social regularities.

- ***Exclusivity***: The positivists' claim that the methods of natural sciences could be extended into the domain of the social sciences including humanities to establish a unified scientific method was also criticized. Positivism excluded the normative questions like beliefs, values, emotions, attitudes and so on. But in reality, much of human behaviour and social arrangements was to be guided by such questions. Hence, it provided a very parochial approach to the study of any domain of social sciences.

- ***Autonomy***: The assertion of positivism that knowledge based on direct observation and verified empirically would yield a scientific discipline that would be objective, neutral and unprejudiced was widely challenged.

Among the social sciences, human geography was the one to adopt the positivist doctrine in a great way as it provided a systematic and scientific approach to the discipline and where reality could be verified. This new paradigm was widely accepted particularly in the fields of urban and economic geography. **Schafer's** paper on exceptionalism opened the door to the domain of (logical) positivism. The geography that developed by adopting the positivist doctrine emphasized on analyzing spatial data and developing spatial theory based on empirically tested mathematical models. However, in the 1970s and 1980s, there was increasing dissatisfaction among the geographers with over emphasis on spatial views and they sought to explore alternative approaches to geographical problems. They argued that human geography employing the spatial view was actually a sort of *'fetishism'* that alienated, diverted and obscured the fundamental social questions. For this reason, even **David Harvey** later deviated from the positivist stand to focus on the question of social distribution.

The critique of positivism in geography mainly emanated from *two* sources:

- Its acceptance of statistical techniques for making inferences about reality; and,
- its acceptance of the assumption of the methodological unification of the sciences.

Regarding the first criticism, **Bennett** in 1985, highlighted the following points:

- Positivism created a false sense of objectivity. The models constructed using statistical techniques that were considered as an effective tool of theory building actually deviated the observer from the observed by giving more prominence to some elements and undermining others. This paved the way for controlling and manipulating society. They were regarded as grossly inadequate for geographical enquiry as well due to non-repeatability of experiments and data.

- By employing quantitative techniques positivism largely eliminated the social and humanistic concerns and reduced humankind as decision-makers or workers to mere passive agents. Such models mostly turned out to be the result of economic determinism.

- Positivism described the existing real situation and thus encouraged status quo in society especially with respect to the distribution of social well-being.

- Since it excluded the normative questions thus it deprived human society of the norms and values based on which it should have been organized. With its overenthusiasm with empirical questions, sometimes it overlooked many good qualitative statements that would have otherwise proved to be effective in describing regional entities.

- Positivism attempted to construct theories with universal acceptability by moving from particular to the general which had reduced validity in real world owing to the spatial character of geographical data. Hence, this resulted into overgeneralization.

- Availability of an extensive and reliable database was an important pre-requisite for the application of statistical techniques in human geography in the absence of which the models or theories developed were supposed to portray an erroneous and distorted picture of reality.

So far as the second criticism was concerned, some geographers like **Peet** or **Slater** derided the positivist-spatial science tradition in geography and even challenged its methodological base as a whole. Their rejection of positivism implied a rejection of the concept of space and hence the rejection of the entire subject matter of geography. They argued that spatial science with geometry as its language was not adequate for addressing geographical problems. Though space was regarded as the central theme of the subject, yet the discipline could not be distinguished from others solely on that ground. The concept of space merely reduced geography to a heterogeneous amalgam of spatial models that yielded no process of understanding or specific theoretical object to geography.

The separate development of physical and human geography also eroded the possibility regarding the unification of the subject. In most situations, human geographers minimized the role of the environment while physical geographers sought to develop theories detached from human and social needs.

The critique of positivism revived the social foundations and responsibilities of social sciences due to which geography went through an anti-positivism and critical revolution in the 1970s. A lot of humanistic approach was proposed to counter and replace the concept of an objective world highlighted upon by the positivist paradigm.

QUANTITATIVE REVOLUTION IN GEOGRAPHY

1. Introduction

In the 1950s and 1960s, a revolutionary change described as "*quantitative revolution*" occurred in the discipline of geography. It replaced the ‘idiographic’ approach based on areal differentiation by ‘nomothetic’ one, which had its roots in the search for models of spatial structure and phenomenon. The *quantitative revolution* led the basis of geography as a spatial science that dealt with the spatial analysis of phenomena that existed on the earth surface. In simpler words, it gave geography a scientific vision through the application of methodology rooted in statistical methods. Some of the elements of positivism, which had previously been not accepted at some point in time, were now accepted open-handedly. In the words of Burton (1963) this school had set out to discover universals, to build models and to establish methods and theoretical bases on which geographical realities could be erected. Traditionally, geography was a discipline that studied and described the surface of the Earth, but in due course of time, its definition and nature have changed. It was now related to providing accurate, systematic, rational descriptions and explanations of the variations in the geographical phenomenon that occurred over the Earth's surface. The most obvious change has occurred due to the *quantitative revolution* that brought changes in the methods and techniques used to explain the geographical phenomenon in a spatial framework.

The movement that led to the occurrence of quantitative revolution in geography was initiated by natural scientists specifically physicists and mathematicians. It expanded and led to change physical sciences followed by biological sciences. By the late 1960s, it became a feature of most of the social sciences. These include economics, psychology, and sociology; though had faint impressions in the disciplines of anthropology or political science, has not occurred in history.

The main objectives of this paradigm in geography were first, to change the narrative character of the subject (geo + graphics) and make it a scientific discipline. The second objective was to explain and interpret the spatial patterns of geographic phenomena in a logical and objective pattern way. The third objective deals with the use of mathematical and statistical techniques; fourthly, to make accurate statements (generalization) about location order; fifthly, to prepare estimates, principles and laws for testing estimates and estimates and forecasts and lastly to provide a sound philosophical and theoretical base to geography, and to make it a scientific discipline.

These objectives lead a number of dichotomies within the discipline apart from the quality dichotomy. Now, these included measurement by instruments versus direct sense-data; rational analysis versus intuitive perception; cold scientific constructs developed in the laboratories versus rich daily sensed – experience from the real world itself; constantly changing phenomena versus discrete cases; nomothetic versus idiographic, to mention a few. If one tries to seek answers to these dichotomies he gets trapped within them and is unable to understand the movement towards quantification in geography. Thus, to avoid this we shall concentrate on how this movement became part of the discipline and slowly engulfed it in such a manner that it led to the spread and growth of scientific method in geography.

2. Quantitative Revolution in Geography

Traditionally, Geography has been a "following" discipline; the main streams of ideas had their roots in other disciplines. The doctrine of environmental determinism was represented in the writings of Semple, Huntington, G. Taylor, and Ratzel (if he can be considered a determinist). They were busy with the idea of a causal relationship and were regularly demanding and looking for "laws". A similar mechanical flavor existed in the works by "Quantifiers". It seems as if geography is re-emerging after it got soaked in ideographic approach, which created a distance between geography and environmental determinism. It seems in some way or other; the quantitative revolution took geography closer to environmental determinism especially as this revolution occurred simultaneously with the upsurge of neo – determinism. The Quantitative Revolution, but natural, was strongly opposed and the dominance of environmental determinism delayed the process of establishment of the scientific basis that the quantifiers wanted to provide.

It was vehemently opposed in the United States as determinism had its strongest base there. Still, new techniques were been used and others were being developed as part of the prevailing

probabilistic trend in contemporary science. In the words of Bronowski (1959) in simpler terms, statistics replace the notion of inevitable effect by probable trends. As the revolution progressed the use and purpose of use of statistical techniques that are quantification became more and more indeterministic.

In geography, the revolution began in the late 1940's and culminated in the period from 1957 to 1960; finally, over in 1963, the year Burton wrote his paper. In between these years, it did gain momentum especially after Ackerman and Schaefer favoured in making geography more theoretical and systematic in nature. Ackerman commented, "although the simplified forms of statistical assistance have been part of geographic distribution analysis in the past; discipline is beginning to move towards more complex statistical methods-a completely logical development'. Burton further commented that both Hartshorne and Spate also agreed on the usage of these techniques in geographical thinking.

The reference of Hartshorne (1959) is being made to his statement, which says, "To raise ... thinking above the scientific knowledge level, it is important to establish generic concepts, which can be implemented with maximum objectivity and accuracy through quantitative measurements which can be subjected to comparisons through the mathematical logic".

Spate (1960) in his paper on "Quantity and Quality in Geography''; published in the Annals of the American Geographers seems somewhat skeptical about quantification in geography. The report of a National Academy of Sciences – National Research Council Committee on 'The Science of Geography' (1965) also discussed the influence of quantitative revolution in geography. They stated that geographers believe that correlation of spatial distributions, considered both statistically and dynamically, maybe the keys to understanding the development of living systems, social structures and environmental changes that occur over the earth surface. In the past progress was slow and gradual as the number of geographers was less while the problems were numerous. Moreover, the methods of analyzing these multi-variate problems were rigorous. It was only recently that systematic concepts and approaches have been adopted to analyze these multifaceted problems.

3. The Path of the Quantitative Revolution in the Discipline of Geography

The roots of the revolution were in the following publications, which had their significant influence on the incidence and growth of quantification in geography. These are – Neuman and Morgenstern's *Theory of Games and Economic Behavior* (1944); Weiner's volume on *Cybernetics* (1948); *Human Behaviour and the Principle of Least Effort* by Zipf (1949) and Stewart's paper entitled *Empirical Mathematical rules Concerning Distribution and Equilibrium of population'* (1947). Stewart's paper needs special mention as he put forward a new way to raise the old geographic questions.

The effect of quantification began to be felt immediately in geography. Rather its rise has been startling in its suddenness. Quantification did increase in geography and one should accept it as it had a valuable role to play. For example, in 1936, John Ker Rose argued in his paper on corn cultivation and climatic conditions that "the methods of relational analysis would be particularly promising tools for geographical investigation." This call was largely ignored. Strahler initiated an excellent petition when he attacked Davis's explanatory and descriptive explanation of geomorphology and supported G. K. Gilbert's dynamic-quantitative systems.

a) Quantitative Revolution in the branches of Geomorphology and Climatology

Strahler claimed that Gilbert's paper was more apter than Davis's work; then what was the reason that it was not accepted as a sign post in geomorphology for future work; rather it has been forgotten and neglected for nearly thirty years. The answer is with Strahler himself who opines that thinks that geomorphology was a part of geography. The physical geographers did not adopt these ideas rather they followed Davis. Some of the prominent followers of Davis include *Douglas Johnson, C A. Cotton, N. M Fenneman,* and *A.K. Lobeck*. Strahler finally states these geographers made "excellent contribution to descriptive and regional geomorphology" and has provided a solid foundation for study in "human geography", but did not lay the basis for scientific study within the geographical thinking. This does not mean that prior to Strahler; geographers were not using quantitative techniques in geomorphology. Quam and Woolridge vehemently criticized his views. Quam (1950) states that mathematical formulae and statistical analysis in geomorphology may result in showing an unrealistic picture of reality that might not be accurate and objective. Similarly, Woolridge (1959) critics Strahler's views and states that although there is the prevalence of a 'new' quasi-mathematical geomorphology; it is inadvisable to use mathematics at a higher level as these are not apt in explaining the geomorphologic phenomenon. He further states that whatever the case may be they will continue to regard W. M. Davis as their founder and would criticize all those who do not agree with the methodology of Davis's interpretations of a different phenomenon occurring over the earth surface. It is not that geomorphologists did not adopt quantification; Strahler did find his support in L. King (1962) who writes that statistical methods are useful for bulk studies and can be well appreciated if used to study complex phenomenon and processes that constitute a large number of variables or indicators. Although few studies in the branch of geomorphology can apply them, they should be used with great precision so that results are not superficial in nature.

Many geomorphologists in addition to Strahler like Chorley, Dury, Mackay and Wolfman, used quantitative methods and it seemed that the practice would spread. In the case of climatology, there is little dispute about the use of quantification. This branch of geography whole-heartedly embraced these new statistical techniques to interpret various climatic phenomena. Examples can be cited from the works of Thornthwaite, Mather and Green, Bryson who have successfully implemented quantitative techniques to seek answers for climate problems; thus silencing their critics.

b) Quantitative Revolution in the branches of Human and Economic Geography

So far, the biggest struggle for approval of quantification has been in human and economic geography. It is not surprising that in view of the possibilist tradition; it is here that the revolution runs against the ideas of independence and the uncertainty of human behavior. Here comparisons with physical sciences are useful. Physicists who work at a microscopic level, with quanta and energy, face similar problems that social scientists face with people. Such parallels when recognized are a reason for happiness and not for disappointment. In order to make a reputable place in human society, social science must get direct results in the form of a prediction science that does not need any kind of control, restriction or regiment the person. A social science that distinguishes random behaviour at the micro-level and is even able to foresee results at this level is nothing but the consequence of quantitative revolution.

Several works can be cited which used statistical techniques in a positive manner. Most interestingly large number of debates took place between scholars that appeared in the literature (Burton, 1963). Some of these are worth mentioning – Garrison's and Nelson debate on Service classification of cities; Reynolds – Garrison's deliberation on the modest use of quantification in geography. The Spate – Berry argument in *Economic Geography* that ends on the agreement that statistics are half of a filled glass, the other half is understanding and interpretations. The list is endless but some of the other debates that need to be mentioned include the contest between Zobler and Mackay on the use of chi-square in regional geography and the dispute of Lukermann and Berry on 'geographic' economic geography.

The deliberations were done through professional magazines, which got them the much-needed attention. The result was the establishment of the Regional Science Association in 1956 that promoted quantification in geography. Moreover, it made quantifiers an essential part of the geographical thinking and giving them appreciation and approving their work part of the geographical academia.

Although most of the literature cites that, the revolution is over, it has remained active in several sub-branches of geography particularly transport, economic, and urban geography. This is evident from the fact that writings with quantitative methods have been regularly published in acclaimed journals in geography, including *Annals of the Association of American Geographers, Geographical Analysis, Environment, and Planning A, The Professional Geographer, Journal of Geographical Systems, Urban Geography,* and many others (Kwan and Schwanen, 2009).

Although quantitative geography is generally "perceived as a relatively static research area," it is actually "a vibrant, intellectually exciting, area in which many new developments are taking place" {Fotheringham, Brunsdon, and Charlton (2000); Clark (2008); Golledge (2008)}. Interestingly, quantification in geography has changed its course in due course of time. It now an ally of critical geographies - for example, the emphasis has shifted from global generalizations to local levels dealing with local relationships in a spatial framework. It has also become sensitive to variables like gender, race, ethnicity, sexuality, and age; and even pays attention to processes which shape individual's spatial behaviour (Kwan and Weber 2003; Poon 2003; Fotheringham 2006).

Quantitative research is still dominant in the fields of transport, economic, and urban geography in the writings of McLafferty and Preston (1997), Rigby and Essletzbichler (1997), Plummer and Taylor (2001), Schwanen, Kwan, and Ren (2008) and Bergmann, Sheppard, and Plummer (2009). In this regard, Kwan and Schwanen (2009) are of opinion that knowledge in statistical methods is essential for decoding and challenging regressive political agendas; often supported by numbers and quantitative analysis. Quantitative geography, when incorporated with a critical sensibility and used suitably, can be a powerful device for encouraging progressive social and political change.

4. The Criticism of Quantification in Geography

The quantitative revolution was initially propounded to make the discipline of geography a scientific discipline where the validity of the knowledge that was generated was justified according to the principles of positivism. Although many geographers like Plummer and Sheppard (2001); Kwan (2004); Fotheringham (2006) have argued that quantitative geography does not necessarily have to be based on the epistemological premises of positivism. Whatever the case may be it is to be understood that when positivist epistemology was adopted, the purpose of the geographic research was to seek universally applicable generalizations. The criticisms became more prominent as critical geographers started questioning the relevance and value of spatial science in the early 1970s. Now quantitative geography was labeled as positivist and empiricist because it was based on the principles of scientific objectivity, value neutrality, and the search for universally applicable generalizations. One of the groups that criticized quantification was the group of feminist geographers that was critical of the tendency to draw conclusions based on the principle of universal causality from inferential statistics (Kwan and Schwanen, 2009). Quantification was also criticized for other reasons. For instance, there were those who thought that this method would mislead geography towards a futile course. Some like Stamp argued that quantifiers were too busy in sharpening their instruments that they forgot the real purpose of the revolution. Few opponents also commented on the suitability of statistical techniques for all kinds of geography. They opine that these techniques were appropriate for some branches and not the entire geographic paraphernalia. Another group condemned this revolution on a note that there was a confusion of ends and means. In the words of Spate (1960), 'it is important to count what can be counted'. Another dichotomy lies in classifying and understanding; classification should never be mistaken for comprehension. Goodall's (1952) point is worth pondering where he states that quantitative methods or statistical techniques are only adjunct to elucidations or descriptions; they can neither provide explanations nor replace them. Therefore, these methods should be observed only as useful tools and not keys to universal knowledge (Spate, 1960).

These criticisms clearly point out that the quantitative methods have some severe limitations, especially when applied to the study of certain kinds of phenomena—for example when the purpose is to uncover the complex gendered, racialized or sexualized experiences of individuals or the socio-spatial construction of identities. However, this does not imply that quantification is not in a position to make valuable contributions in the field of geography. The difference lies in the time period if we talk of the 1950s or 1960s maybe this was not possible but in contemporary geographic research, it is possible to reconnect the critical geographies with quantification. Another point of deliberation is that within the discipline of geography several subfields, like transport geography, are historically more quantitative in nature than others; this happened because of the influence of allied fields such as civil engineering and neoclassical economics (Kwan and Schwanen, 2009).

It can be said that the revolution had an early demise; it means that the purpose of the revolution was achieved or not. If seen from the point of view of Burton (1960) its basic purpose was to make geography more scientific and to develop a body of theory. Discontent with the idiographic approach in geography is the root of quantitative revolution; the development of theoretical and model-building geography with a nomothetic approach was the expected result. The basic rationale was to develop scientific method; to develop the theory and to test the theory with the prediction for which the logic of mathematics is the best tool available.

5. Conclusions

The use of statistical or quantitative techniques is one of the most suitable methods for the development of theory in geography. The revolution can never be over until it is able to seek answers and aid the theoretical development of the discipline. Moreover, theory development and its testing are the only ways of creating new knowledge and subsequently verifying it. Models have just formalized ways of descriptions that an author has visualized and represented through his arguments and justifications. In geography, quantification brought this revolution where the ideographic base was replaced by theory building in a nomothetic approach.

Geographers started developing theories and created '*new*' geography that focused on the philosophy as well as methods. These scholars were of the view that mere description, mere quantification, and mere abstraction were valid to a certain extent; but repeated use of these methods makes one an obscurantist. *Theoretical geography* got its philosophical base in Bunge's monograph published in 1962, which identified geometry as the mathematics of space and hence made spatial science the language of new geography. Harvey's *Explanation in Geography* (1969) provided a more inclusive channel for the methods and philosophy of new geography. Apart from these scholars, the Department of Geography at Lund University, Sweden became a centre for quantitative and theoretical geography under the leadership of Hagerstrand and Morill. Hagerstrand although based in Seattle provided an academic support to the geographers working in this field at the Lund University. To conclude, whichever method one, the purpose of geography is to seek answers to questions pertaining to problems of quantity and value. Most of our experiences are qualitative in nature and when everything is, reduce to numbers; some essential attributes are lost (Huxley, 1951). Thus one needs to maintain balance as still new worlds are to be conquered and new contributions to be made (Bansal n.d.).

SCHAEFER –HARTSHORNE DEBATE

I. Introduction: The debate on the methodology of pursuing Geographical studies that ensured between the two geographers F K Schaefer and Richard Hartshorne is one of the most stimulating and academic exchanges the subject of geography has witnessed. This debate began with F K Schaefer's paper titled "Exceptionalism in Geography: A Methodological Examination" published in the Annals of the Association of American Geographers in September 1953. Schaefer's articles were published posthumously, and the debate continues as Richard Hartshorne published his comment (caveat) in the same journal in 1954, followed by an elaborate article titled "Exceptionalism in Geography Re-examined" published in the same journal in September 1955. Richard Hartshorne also published another article in the same journal in June 1958 titled "The Concept of Geography as a Science of Space, from Kant and Humboldt to Hettner." Drawing from the research articles mentioned above the debate on the methodological position of Geography that arose in the academic circles of the United States in the 1950s may be charted out.

II. Exceptionalism in Geography: In the early Nineteen fifties, the geographers in the United States were influenced by the regional paradigm as the chosen methodological frame for pursuing geographical research. The most celebrated proponent of the Regional paradigm was Richard Hartshorne also from the United States and Hartshorne's seminal monograph named "The Nature of Geography" which appeared in the Annals and later was published in the form of a monograph by the Association of American Geographers. Though Hartshorne's paradigm of regional geography was well received, there arose disillusionment and discontent in accepting regional paradigm as the most accepted way of doing geography among certain sections of geographers in the United States.

F K Schaefer was a trained as an economist and later pursued Geography. Schaefer was associated with the Department of Geography at the University of Iowa and he had migrated to the United States to escape the Nazi persecution that was taking place in the then Germany.

F K Schaefer took it upon himself to clarify the position of Geography within the broad arrangement of various schemes of knowledge systems; the sciences and the social sciences. His motive was to refute the pre-eminence given to Regional Paradigm of Geography as the only mode of conducting geographical research and for that, he re-interpreted the works of scholars whom Richard Hartshorne had used in his monograph.

F K Schaefer's rebuttal to the Hartshornean tradition of Regional paradigm came in the form of a research article published posthumously in 1953 by the Annals of the Association of American Geographers. It is in this paper that he takes up the cause of systematic geography while rejecting the claim of Exceptionalism of Geography.

The claim for Exceptionalism in geography emerged from the notion that the core focus of geography should be areal differentiation. Geography should be studied as a science that explains the realities of areal differentiation that appears on the surface of the earth. The differences between these regions need the attention of geographers as well as the combination of all aspects present on those spaces and how they are different from other places. In this manner, Geographical research would be able to put forward an "accurate, orderly and rational description and interpretation of the variable character of the earth's surface."

The aim of geographical research according to Hartshorne is to gain a complete understanding of the areal differentiation of the earth. For that, geography must focus on the synthesis of all phenomenon occurring in a particular region. The region too would get its identity from the synthesis of the various phenomenon present in the region.

Hartshorne also leans on the Kantian tradition and draws an analogy between history and geography, saying that as history divides time into sections so does Geography divides space into sections. Geography thus studies the world and describes and interprets different regions as they appear in a particular time. What other disciplines study as a heterogeneous phenomenon, geography studies that entire phenomenon in combination. By doing this geography puts together the aspects which other disciplines study in isolation.

For looking at the phenomenon in combination, geography must use descriptive mode of analysis For Hartshorne the raison d'etre of Geography is the study of areal differentiation, so it can be best expressed through regional geography.

Regions are categorized into 'formal regions' and 'functional regions'. Formal regions are those where the entire region has a homogeneous phenomenon and Functional regions are those where the unity of the region depends on the node which controls the flow of the phenomenon across space thus carving out a functional region.

Regional paradigm this acquired the status of the most feasible methodology of conducting geographical research; as regional geography brought together various aspects of spatial character which are dealt separately in topical geography. This method of bringing together of the phenomenon and their interaction with each other to give a special character to a region was analyzed in geographical research. Even within topical specializations, the study of regions had an important role to offer. If we take the example of relating Economics with Economic Geography and applying the theories and laws of Economics to Geography, the focus of the Geographer as distinguished from the Economist would be his application of the economic theories over and within regions.

Hartshorne also elicited the enquiry of 'generic problems' rather than 'causal relationships' while undertaking geographical research. He was keener on the functional representation of phenomenon on the earth's surface rather than how each phenomenon leaves its imprint on the surface of the earth. With the focus on the organization and arrangement of the phenomenon on the surface of the earth as the primary research objective of Geography, enquiring into the causality of the assemblage of phenomenon took a backseat.

III. Challenging the Claim for Exceptionalism in Geography

The descriptive methodology for doing Geographical Research met with resistance among those practitioners of the discipline who were of the opinion that geography should also be a law making discipline and not remain as a discipline dealing with only descriptive methods.

F K Schaefer became the voice of the discontented Geographers and took it upon himself to write an article challenging the 'Exceptionalist' claim for Geography and urging geographers to adopt the methods used by Positivist School of research which may help elevate the discipline to the status of a law making social science.

Schaefer's thoroughly well-researched article attempts to dispel the confusion on the nature of the discipline of geography is as attempted by earlier geographers whom Hartshorne referred to while upholding regionalism as the method to be followed by geographers. Schaefer contested treating all historical accounts of the development of the discipline as a method of teaching the same. Methodology deals with the position of the discipline vis a vis the systems of knowledge and methodology undergo constant evolution and change with the passage of time. The scope and content of geography as viewed by geographers have been apologetic and is always eager to justify the very existence of the discipline by offering it the status of ‘integrating science’. It has been agreed that the progress of geography has been slower compared to other social sciences like economics, but yet it is undoubtedly an important discipline.

In the nineteenth century, the growth of the natural sciences elicited the fact that mere descriptive study would not elevate a discipline to the status of science. The laws that geography may come up with are mostly about the formation of patterns which are spatial in nature.

Humboldt and Ritter too accepted the fact that all spatial patterns are governed by laws including those created by natural phenomenon and man.

Schaefer looks into the works of Viktor Kraft that geography is a science trying to discover laws that are related to the surface of the earth. With this Schaefer brings in the case for systematic geography where spatial relations of two or more phenomenon is studied over the entire earth’s surface to arrive at a generalization or law. He blames the lack of clarity on the relative role and importance of regional and systematic geography to the differential preference given to the two methods at different phases of the development of the discipline. In the beginning of the nineteenth century, Physical Geography applied systematic method while at the beginning of the twentieth century the shift occurred towards research in human and social geography and since not much of social science-based laws had developed during that period, the systematic method floundered. Generalization was found unpractical and the practitioners of geography aligned themselves with the opponents of the scientific method.

Schaefer then quotes Hettner and dispels misconception on his use of the word ‘dualism’ for Geography. According to Schaefer, dualism was not used to emphasize that Geography is a methodologically unique discipline, nor is the complexity a regional geographer face is facile.

Schaefer also delves into the historical roots of this uniqueness of Geography which he terms as ‘exceptionalism’ here. He calls Immanuel Kant as the father of exceptionalism but terms him as a poor geographer of the eighteenth century when compared to Varenius. Kant claimed Exceptionalism not only for Geography but also for History. Kant categorizes History and Geography both as descriptions with History is a description according to time while Geography is that of space. This distinction according to Schaefer is untenable as no systematic discipline can ignore the coordinates of time and space. Secondly, simply because Kant says that both the disciplines are descriptive, that there exist no laws in the two disciplines. Humboldt too in his celebrated publication ‘Kosmos’ explains that all sciences search for laws and are ‘nomothetic’. Schaefer also accuses use of Hettener’s great prestige to create more confusion around the exceptionalist claim by drawing spurious similarities between History and Geography. It is through Hartshorne that the American geographers have been acquainted with Hettener’s work but Schaefer feels Hettener’s work can equally be supportive of a nomothetic method for geography. But the impact of Exceptionalism on geography has been deep and it needs the dispelling of the misconceptions around it. One way Geography s different from other sciences is that sciences progress towards the formulation of laws, but geography is essentially morphological so often the time factor is ignored. The breakdown of ‘causality’ in geographical research led to its failure as a discipline. The morphological character is often represented by maps as a tool and maps are selective since only certain features that are of interest to the researcher is mapped and also help in correlating phenomenon spatially. This method is termed as comparative geography but is, in reality, systematic geography.

Quoting Palander, Schaefer also tells that the notion of regions explains nothing and is no substitute for a spatial law. By rejecting the formulation of scientific laws, the possibility of prediction is compromised by a discipline. Laws of geography can be categorized into three classes. Laws of physical geography fall in the first category, the laws of economic geography in the second category and the non-morphological laws come under the third category. Mature social sciences look for ‘process laws’. If geographers do not pursue systematic method and strive to formulate laws, it might lead to geographical isolationism among other disciplines as the search for laws can occur only with cooperation with other social sciences. Schaefer recommends pursuing systematic method for enhancing the prospects of the discipline. He casts his doubts on the future of geographical research if geographers restrict themselves to search for regions and not align themselves with other systematic sciences.

IV. Exceptionalism in Geography– Re-examined

Richard Hartshorne expressed his reaction to Schaefer’s paper in the form of his articles at the same journal Annals of the Association of the American Geographers in the years 1955 and 1958. In 1954 to a brief comment was published by him which was a caveat to the geographers to critically look into Schaefer’s paper and that he would publish the clarifications on his work ‘The Nature of Geography’.

In his article titled ‘Exceptionalism in Geography – Re-examined’ published in 1955, calls it a correction to the false representations and accusations made by F K Schaefer in the 1953 article. He also admits, that his paper does not deal with the only negative purpose of claiming all that Schaefer has written is false but he too in the process of writing the clarifications learned a lot on the type of methodology suitable for geographical research. As Schaefer noted that methodological discussions are dialectical, knowledge may be derived from mutual discussion and debate. He also notes down the rules for Methodological Writing in this paper.

The need for Re-examination of the Exceptionalist Claim for geography arises according to Hartshorne since the article by Schaefer remains in the Annals for future students to get misguided, it's important for the clarifications. Hartshorne writes that methodological studies on Geography have been most influential in Germany who have altered they view from the earlier Geographers. He names Hettener; and also Kraft and Humboldt saying that the writing of Humboldt mentioned by Schaefer does not have any methodological topics and Kraft’s essay has been mentioned yet not considered. Most other important authors like Richtofen, Mackinder, Valaux or Sauer have not been mentioned by Schaefer.

The students thus may be led to ‘almost devastating’ conclusions. He also claims that the basic purpose of the essay written by Schaefer is unclear to him and leads to another major issue of ‘exceptionalism’ which in reality does not exist. The major thrust of the essay is that geography must be a science, a science that searches for laws and no new arguments have been presented.

Texts which reject Scientific Determinism like that of Hettener has been attacked as well as Hartshorne’s own ‘The Nature of Geography’. The article exposes ‘defective scholarship’ assertions without supporting evidence. By organizing the paragraphs and arguments in a rather confusing way the readers are brought into having the same opinion as for the critic. The major targets of Schaefer’s essay are the writings of Hettener and Hartshorne. Certain views about a few authors have also been proven as wrong directly or by implication. Even false translations have been used in the essay to present false impressions of the original work of the writers. The falsification and fabrication of the facts had been deliberately made and writings have been distorted to generate an opposing opinion of the readers. Hartshorne’s sharper critique of Schaefer’s work came in the form of his 1959 monograph titled ‘Perspective on the Nature of Geography.

In this monograph, Hartshorne defines the discipline of Geography as one which “seeks to describe and interpret the variable character of from place to place of the earth as the world of man”. According to Hartshorne, the physical and the human aspects of an area need not be dealt with separately but as an assemblage of the interaction between the human aspects with the physical factors. The reliance on the previous paradigm of Environmental Determinism shapes the understanding of the way the human aspects in a region would adapt and respond to the physical aspects of the region. He also considers the division of Geography into Physical and Human aspects as unfortunate. By splitting the subject matter into two; geographers lose out on integrating both the aspects (physical and human) while undertaking research.

Hartshorne also clarified the difference between ‘exploratory description’ and ‘explanatory description’ while adopting the descriptive science methodology in research. Hartshorne warns that explanatory description of the former geographies or assemblages should not be given pre-eminence while doing geographical research. In his account, Historical geography should be undertaken not to ascertain origins of the phenomenon, but to offer a better understanding of the present phenomenon in the light of the past.

Hartshorne, however, accepted that there is a possibility of collaboration across the spectrum of methodologies of regional and systematic geography. He said that those researchers who are engaged in organizing areal differentiation of elementary complexes in large scale like the entire world can be linked with those researchers who are involved in deciphering complex relationships of a phenomenon over small- scale regions. He calls them topical studies and regional studies respectively.

Geographical studies apply both the methods while conducting research, and no one method can be called more important than the other. Integration of both the methods is also possible. Schaefer tried to raise another important issue regarding Geography being Idiographic or Nomothetic discipline. Schaefer asked whether geography should formulate scientific laws or indulge in the descriptive analysis of individual cases. Hartshorne was of the opinion that since scientific laws require abundance of cases to prove itself, and that geographical research limits itself to limited numbers of observations; so geographers should refrain themselves from formulating scientific research. He also said that scientific laws are best tested in the laboratories and geographers are usually not in the position to conduct their experiments in the laboratory. Hartshorne also opined that the interpretation of the scientific laws requires training in systematic sciences which the geographers lack and may not be in the position to offer the right interpretation of the phenomenon. Hartshorne also said that scientific laws also rely on some form of determinism. When conducting research on human subjects such determinism appears ‘inappropriate’ as the difference in human ideas lead to landscape variations. For the reasons mentioned above, Hartshorne suggests that geographers should stay away from searching laws in geography.

Hartshorne also reiterated his support to Hettener’s classification of geography as chorological science and history as chronological science bringing back Kantian ideas of history divides knowledge through time and geography through space.

Geographers in their research have also tried to integrate topical and regional aspects within demarcated areas of research thereby often integrating both the methodologies. Hartshorne’s ideas regarding Geography were based on how Geography as a research area was shaping up during his times in the United States. Schaefer, on the contrary, wrote about what geography was not. By that Schaefer wanted geography to incorporate aspects of research the discipline was neglecting and how Geography could claim its lawful place within the realm of social science research.

V. Conclusion

The debate that ensued between F K Schaefer and R Hartshorne was one of the rare exchanges one finds in the discussion on the methodological aspect to be adopted in the discipline of Geography. During the mid-twentieth century, many schools of Geography in the United States were of tired of the regional paradigm and were willing to adopt a more rigorous paradigm for doing research. With this debate spreading across academia, the gradual change towards topical specialism as a methodology was witnessed in geography (Basu n.d.).

SYSTEMS APPROACH AND SYSTEM ANALYSIS

I. Introduction

The word system has been derived from the Greek word “system” which means a set of rules that govern structure and behaviours. In other words, the system is termed as a unified whole (working body) which consists of interdependently functioning elements. The element is very basic part of a unified whole. For example, the human body is a biological system involving various elements (parts) like cells, tissues, blood, bones, and muscles. These elements (parts) are functioning interdependently. Likewise, the Earth itself is the largest system which is made of lithosphere, hydrosphere, atmosphere, and biosphere. The biosphere is the largest ecosystem made of interconnected sub-systems (both terrestrial and aquatic ecosystems) viz., forest, grassland, desert, ocean, lake, pond etc. These systems vary greatly in size and scale ranging from microscopic to micro, meso and macro. For instance, biosphere forms an ecosystem of macro size and rivulet can form a micro one.

A. System Approach to Geography

The concept of the system approach has been fundamentally derived from the general systems approach or theory. A biologist named Ludwig von Bertalanffy's seminal paper on open systems is attributed as a seedling for the rise of the system movement. He has published various papers on a system approach to biology between the 1920s and 1950s. His papers aimed at giving account for the key distinction between the organismic systems of biology and the closed systems of conventional physics and understanding common laws that govern the life of organisms. Through his general system approach, he comprehended intrinsic unification of different streams of sciences and fusion between science and environment. For von Bertalanffy, the main propositions of general systems approach or theory were:

1. Isomorphisms between the mathematical structures in different scientific disciplines could integrate and unify the sciences;
2. Open systems require consideration of the flow of energy, matter, and information between the system and its environment;
3. Within open systems, the same final system state may be reached from different initial conditions and by different paths – open systems exhibit equifinality;
4. Teleological behaviour directed towards a final state or goal is a legitimate phenomenon for systemic scientific inquiry;
5. A scientific theory of organization is required, to account for wholeness, growth, differentiation, hierarchical order, dominance, control and competition; and
6. GST could provide a basis for the education of scientific generalists

As per the above-mentioned propositions, a general system is a unified whole of elements bound together by specific linkages. It is higher order generalization of a multiplicity of systems, their complex structures, and functions. This is an analytical framework to unify various sciences. It has a self-sustaining mechanism. Such systems may be open or closed and may change over the period of time. Most general systems, however, are open. As discussed earlier the earth is an open system in which there are inputs, outputs, and energy flow through a variety of mechanisms. The linkages, or connections, that bind entities together into a system, are paths through which matter, energy, ideas, and people pass from one element to another.

As very early, Ludwig von Bertalanffy had realized the importance of system approach to non-biological science. Over the period of time, geographers introduced system approach to various geographical studies. R. J. Chorley, Leopold and Landbein, Wolderberg and Berry were some prominent geographers who made its application possible in geography. In fact, it was R. J. Chorley, who become the first geographer brought about this approach or theory to geography. His paper "Geomorphology and General system theory" (1962) focused on the mainly application of the concept of open and closed system in Geomorphology. In addition, Leopold and Landbein applied entropy and steady state in the study of the fluvial system. Berry developed a basis for the study of the city as a system within the system of cities in spatial form by using the two concepts viz., organization and information. Wolderberg and Berry have applied system concept to analyze central place and river system. Curry has tried to use this concept to the spatial location of settlement. Even new age geographers solve numerous spatial problems through a general systems approach in both human and physical geography. They use the system as analytical models to understand and explain spatial patterns and interactions. Human geographers, for example, use the system model to examine human migration patterns, the diffusion of ideas, and the spread of information. Moreover, researchers in physical geography trust this approach in understanding natural set up in which physical system operates. In short, researchers in both physical and human geography are interested in identifying, explaining, and predicting flows in human and physical systems. They also search, identify, describe, and explain cycles and patterns of geographical phenomena in different branches of geography.

In other words, the systems approach is used in a variety of applied branches of geography viz., land-use planning, natural resource management, watershed management, regional planning, management of pollution (water, air, soil, sound), environmental management, climate change etc. These areas involve the study of elements and sub-system of general environment like quantitative, qualitative, behavioural, socio-economic, and political subsystems. The qualitative subsystem encompasses finite space including urban, rural, empty or filled places, psychological spaces, and their various-use. It also includes a variety of concepts like capability, carrying capacity and stability. The economic subsystem comprises decision-making processes based on well-tested economic theories whereas citizens, governments, civil administration, and civil societies form the political subsystem, which plays a vital role in geopolitics of a state (nation) being studied in political geography. The behavioural subsystem consists of attitudes, values, beliefs, customs, and traditions, which are integral parts of behavioural geography. A general systems model is a composite one in which physical and socio-economic variables are intricately linked. Some of the variables may be measured quantitatively and some may not. The systems approach involves a number of relationships (links) between variables (elements). Now geographers use statistical tools like multi-variable analysis, principal components analysis (PCA), probability theory, Chi-Square and Gini-coefficient to analyze data on geographical variables explaining interlinks between them in a system.

II. Systems Analysis

A. Basic Elements

System, as discussed earlier, is a unified whole or working body, which consists of interdependently functioning elements. There are multiple variables or elements that form a system. A system forming basic elements is as follows:

1. **Inputs and Outputs:** Every system requires a regular flow of inputs for producing an amount of output. Without inputs, no outputs can be produced in a system. For example, a fixed agricultural output needs a certain amount of inputs of seeds, water, fertilizers, labor etc. in the assumed constant external environment.

2. **Processors:** It involves the actual transformation of input into output. It is the operational component of a system. Processors may process the input totally or partially, depending upon nature, amount, and requirement of the output.

3. **Control:** This is an important element, which guides and controls the system. It is basically the decision-making subsystem that governs the pattern of activities like processing input and producing output.

4. **Feedback:** Feedback is an indication for characteristics, amount of produced output against the set parameters and standard. Feedback is conducted on the principle of

cybernetics which comprises communication and control. Feedback is of two types., viz., positive or negative. It is good to have positive feedback, which strengthens system's performance. Negative feedback gives the wrong signal if the system is not functioning well and it also makes available the controller with information on actions required for its correction.

5. **Environment:** Every system is operating in a unique environment. It is a broader framework often called "supra-system", Which affect operating system and determining routes and rules of its functioning.

6. **Boundaries and Interfaces:** A system has delineated boundaries or limits through which it identifies its components, processes, and interrelationships when it interfaces with another system or with its environment.

B. Components of a System

A set of elements form component and a set of components form a system. All systems of varying scales are having three basic components as follows:

1. A set of elements

2. A set of functioning links

3. A set of links between system and external environment

All systems have internal and external environments. The external environment influences the internal environment of the system. Some systems are close and some are open. A closed system can easily be created in science laboratories. In other words, the closed system exists mainly in a controlled environment like chemical labs. For example, chemists conduct chemical tests in their labs but such kind of total control is not possible in an open environment like agricultural or forest land, but in open environments like agricultural land, some elements can be partially controlled. Such partially controlled environments are of great importance for semi-scientific experiments aiming at socio-economic development. This is to be understood with this example. If farmers want to know impacts of certain inputs on a crop production. Impacts of inputs like fertilizers, pesticides, high-yield seeds, labour on crop production can be known by controlled and regulating inputs in a farm. With above discussion, the basic characteristics of a system can be inferred as follows:

1. System is a well organized and an integrated ideal body;
2. Systems have multiple elements and components;
3. The components in a system are interdependently functioning;
4. Systems have a structure and pattern of behaviour;
5. Systems have a boundary and interfaces;
6. Systems are at balancing and enduring state;
7. Systems affect and are affected by their internal and external environment;
8. Systems exhibit feedback;
9. Systems are either closed and open and
10. Predominant systems in the environment are open-ended.

C. Essential Features

Above mentioned basic characteristics exhibit some interlinked essential features of a system as follows:

1. Environment of a System

2. The Behaviour of a System

3. The State of a System

4. Organization and Information in System

5. Structure of a System

1. Environment of a System

The environment of a system is supra-structure in which system operates. There is countless system and their sub-systems are working in our environment. The environment is classified as physical environment, socio-cultural environment, political environment. Socio-cultural environment and political environments operate in their physical environment. In fact, each system has its own physical, socio-economic environment that affects the performance of that system positively or negatively. For example, farmland is a system, its agricultural productivity depends upon its physical, socio-economic environment. A system may have the internal and external environment. For example, the productivity of an industry as a system is affected by both internal and external environment.

2. The Behaviour of a System

It refers to all dynamic activities of a system like the introduction of new stimuli, flows, and responses, inputs, and outputs etc. It studies the flow of energy between the elements of a system and between a system and another system. Functions within a system are called internal behaviour and outside a system is external behaviour. Internal and external behaviours are interlinked. For example, an element of a system is the part of the external environment, change in external environment will bring some behavioural change in one element and change in one element will affect all inter-connected elements of that system. Such behavioural change can be explained by its flow from input (simple stimulus) to the output (response).

The input-output analysis in economics is a popular example for explaining the flow of behavioural change. Increase in final demands (e.g. derived from exports, home consumption, or another way around) is working as stimuli to rise in final outputs in various sectors in an Indian economy (a system).

3. The State of a System

Each system wants to achieve its state of equilibrium. A slight change in one element of system disturbs its equilibrium. A disturbed system may experience morphogenesis to gain a complex level of equilibrium. 'Equilibrium' denotes maintaining a kind of balance in a system. Equilibrium is two categories, viz. stable and dynamic. Further, the stable equilibrium includes both homeostasis and steady states. A homeostatic system shows always activity, but it does not alter the balance between the system's components. A system in a steady state is also stable, but it may change in an orderly way. Dynamic equilibrium represents a process by which a slight disturbance causes a constant change in throughout the system.

4. Organization and Information in System

Normally system is well organized, that's why one can predict the possible amount of change in all set of elements of a system if a change occurred to an element is known but such prediction is not possible if the system is not organized. **'**Information' is considered as 'the measure of the amount of organization' (as opposed to randomness) in the system. Good information means well-organized system. In addition, the word 'entropy' and 'negentropy' are associated with organization and information. Changing level of energy in a system creates a disorder, hence disorganizing the system. In other words, entropy (a measure of unavailable energy) is regarded to be a measure of disorder or disorganization of a system. In the contrary, the negative entropy or negentropy, on the other hand, is a measure of order in a system. The close system may have highest entropy hence making a system inactive. While interacting with the environment, it is good for the open system to have the optimum level of entropy creating a more complex system. As entropy brings disorder and negentropy brings order, these concepts can be used in different branches of geography like an ecosystem, river system, and socio-economic systems.

5. Structure of a System

Structure of a system depends on how element and components of a system are arranged and interlinked. Therefore, the structure may vary in terms of its shape and size. The structure could be hierarchal or parallel. For example, in Walter Christaller's Central Place Theory, settlements are hierarchal arranged in an urban system. Large urban cities in few number lie at the top of the order and small cities in large number at the bottom.

D. Common Relationships

The links of elements shape components and structure of a system. A different pattern of links between elements forms a variety of relationships, some of the most common relationships are being illustrated through Figure no. 3, 4, 5, 6 to 7.

1. **Cause and Effect Relationship:** This is the simplest relationship which is also called 'Series' relationship in which elements are connected by an irreversible link.

For instance, rainfall affects the rate of soil erosion but soil erosion directly does not have an effect on rainfall.

2. **Parallel Relationship:** When two elements affect third element making relationship called parallel one. For Example, rainfall and temperature affect vegetation and vegetation, in turn, directly or indirectly affects the amount of rainfall and local temperature.

3. **Feedback Relationship:** This is newly introduced relationship into analytical structures. Two elements get mutually affected. For example, farmers grow pulses (leguminous plants) which enriches nitrogenous fertilizer in the soil and in turn, production of pulses increases because of enriched soil.

4. **Simple Compound Relationship**: In a simple compound relationship, components are modified by itself and influenced by a set of other external components. Both processes operate simultaneously. For example, industries in India are removing their old technologies and adapting new foreign technologies to increase low-cost production in the competitive global market.

5. Complex Compound Relationship

This is the most complex relationship of elements in which elements of the internal and external environment is mutually affected and influences each other. In our environment, all real systems have complex compound relationships amongst their element and complement. Our ecosystem is the best example of a complex compound relationship, which is very difficult to interpret correctly.

E. Classification of Systems

On the basis of above-mentioned relationships, salient features, and characteristics, Systems can be classified as homeostatic, dynamic, self-regulatory, adaptive, controlled systems etc.

1. **Homeostatic System**: A constant balance maintained in a system is named as a homeostatic system. Such system by its constant operating environment restores its equilibrium or steady-state behaviour if it faces some external interventions. As per its nature, it resists an alternation caused by internal environment but if it faces new change, its processor restores previous equilibrium or steady-state. There are innumerable homeostatic systems in our environment. For example, the human body is a homeostatic system that maintains its equilibrium in its temperature at about 98.2 degrees Celsius. Temporarily, it might change but body againrestores equilibrium in its temperature.

Likewise, innumerable geographical systems operating in our environment are known as a homeostatic system which maintains equilibrium or steady-state. The geomorphic cycle of erosion is the homeostatic system. In the cycle of erosion, if any element like the amount of water, slope, suspended particles etc. changes, the entire system gets affected but with certain changes, cycle maintains steady-state.

2. **Adaptive System**: It is a system which has adaptive capacity to changing external environment. It's some characteristics are similar to the homeostatic system. This system sustains a constant operating environment to achieve the desired state which has been emerged because of certain change in external environment. For example, our socio- economic systems are becoming adaptive to climate change. New technologies are introduced in agriculture and energy sector in wake of climate change. The direction of the adaptive system depends upon the feedback, it is getting in form of increased or decreased productivity.

3. **Dynamic System**: It is different from both homeostatic and adaptive systems which experience some change over the period of time in achieving steady or desired states. The dynamic System shows a chain of continuous changes along with a line behaviour in the entire system over the period of time. For example, the vicious cycle of poverty and cumulative causation as Economic growth models.

4. **Controlled System**: It is a system in which elements or inputs can be regulated to achieve goals (results) for socio-economic development. Normally such kind of system lies in the close environment like laboratories. For example, scientists, doctors, and chemists conduct experiments to assess the impacts of certain chemicals as medicines on animals or human bodies in a controlled environment. In a study of system engineering and cybernetics (the study of communication and control mechanisms in machines and living beings), partially controlled systems are of great importance. Even in geographical subject matters like resource management, regional and economic planning, partially controlled systems can be created and applied. Example, economically backward region can be developed by pushing huge investment in infrastructures hence creating employment opportunities for local people but such environments cannot be completely controlled. Therefore, partially controlled environments are of great importance for human well-being.

In conclusion system approach and its analysis offers a simplified theoretical and conceptual framework to study the subject matters of geography like study of landforms, river system, ecosystem, regional and economic planning and social and economic development etc. Geographers can apply this approach in all three stages of research viz., descriptive, analytical (explanation and seeking governing natural laws and undegrading orders in the real world) and predictive (how existing orders are likely to change in future?).

III . Advantages and Disadvantages

Application of system approach and analysis in geographical studied has various pros and cons. It discloses inherent information on current state, structures and dynamic behaviours of various geographical systems. Our open existing environment is so complex that it goes beyond our understanding. Therefore, system approach simplifies existing environment in order to make it easy for students to understand. In words, it is a technical tool to comprehend interaction between elements of any complex geographical structure in simplified ways. It also helps us to develop a variety of abstract geographical theories. More importantly, its mathematical languages like geometry and probability theory are widely used in solving numerous geographical problems like rising pollution and prediction of climate change and understanding affecting factors. Despite these advantages, this approach is criticised because its overemphasis on positivism and quantification social science (quantitative revolutions) and avoiding normative values (beliefs, attitudes, desires, hopes, fears). Nevertheless, system approach is still relevant in geography (Jiwan n.d.).

Unit - 6: DEVELOPMENT OF HUMANISTIC GEOGRAPHY

Humankind as an agent of change on the earth's surface was first identified by **Comte de Buffon** as early as in the 18th century. Inspired by his ideas, **Immanuel Kant** developed his physical geography that was essentially *'anthropocentric'* in nature and content. According to Kant, physical geography not only included the features visible on the earth's surface created by natural processes but also by human actions. Kant was also of the opinion that empirical knowledge could be obtained in two ways—either (i) through pure reason, or (ii) through the senses. Senses again could be divided into—(i) inner senses and, the (ii) outer senses. The world as perceived by the inner senses was the ***seele*** (soul) or ***mensch*** (man) while as perceived by the outer senses, was the Nature. The concept of Kant's anthropocentric geography was subsequently adopted by **Carl Ritter**. In his famous ***'Erdkunde'***, he asserted that the central theme of geography was the element of reciprocity that is believed to have existed between the natural phenomena and humanity. Subsequently, **Friedrich Ratzel** in his ***'Anthropogeographie'*** set a framework for the systematic study of human geography and thus set a new trend in the subject. Prior to him, systematic geography only involved physical geography and, human geography was mainly confined within regional studies. His anthropogeographie was essentially a reflection of the Darwinian viewpoints and emphasized on the concept of natural selection that was used in the natural sciences.

The human approach in geography was greatly popularized by the French geographer **Paul Vidal de la Blache** in 1899 with his introduction of a new dimension to the possibilistic philosophy. Blache may rightly be called the *father of modern human geography*. He advocated ***'genre de vie'***, a concept akin to human culture, inherited and developed over time to convert natural ***'possibilities'*** into elements of fulfilment. Nature was conceived as a mere adviser and

humanity, an active force of change. Blache's possibilist philosophy was carried forward by **Jean Brunhes** throughout France and other parts of the globe. His main emphasis was on the exploitation of the earth by humankind for satiating human needs and desires.

In fact, it was the French historian, **Lucien Febvre**, who is actually credited for coining the term ***'possibilism.'*** In his ***'Geographical Introduction to History'*** (1922), Febvre accorded to the Vidalienne tradition of possibilism. He put forward that humankind emerged as a powerful agent of modifying the earth's surface through centuries of his accumulated labour and decision- making. In 1924, American geographer **Carl O. Sauer** propounded his ***'landscape paradigm'*** in which he highlighted on humans as agent of *'fashioning'* the natural landscape.

The discipline of geography underwent several paradigm shifts and revolutions over time. The 1920s witnessed the revival of the positivist philosophy after. The concept was however, introduced in the 1830s by **August Comte**. The aftermath was a theoretical and quantitative revolution in geography in the 1950s. **Schafer's** critique of Kant's exceptionalism and the introduction of his ***'spatial organization paradigm'*** opened the door for such revolution. However, in the 1970s there was yet another revolution in geography which was essentially anti-positivist in nature. It came to be known as the ***'critical revolution'*** as its origin was rooted in the criticism against the **positivist-quantitative-spatial tradition of geography**. The effort was on replacing the quantitative methods with a variety of humanistic approaches. This was supposed to ascribe a pivotal role to humankind in the subject particularly to ***'human awareness, human consciousness and human creativity'*** and freed human beings from geometric determinism. Thus, the modern humanistic geography was mainly an outcome of the growing dissatisfaction against the quantitative revolution.

Effort was made to revive the *'normative statements'* of values, attitudes, beliefs and so on. It aimed at ***'verstehn'***, that is, understanding humankind within the surrounding environment in which humankind by using his rationality could improvise on the conditions of their lives.

The proponents of humanistic geography asserted that, humanistic geography should not be considered as an earth science in its scope and content. Instead of viewing geography as the study of the earth, it treated geography as the study of the earth as the home of humankind.

Hence, the main focus was on how humans perceived the place they inhabited through their thought processes, consciousness and experiences.

Humanistic geography sought to be more than a mere critical philosophy. Therefore, **Anne Buttimer** in 1978 attempted to resuscitate the Vidalienne tradition and asserted that any spatial unit should be studied from a *local perspective* (similar to Blache's concept of *'pays'*) and with a *historical approach*. This was possible because some affinity was discovered between Vidal's *le geographie humaine* and humanistic geography. But, there were grounds of departure between the two. Firstly, Blache considered human geography as a natural science and, many of Blache's work contained the elements of functionalism which the humanistic geographers renounced.

Humanistic geography also contained elements of ***neo-Kantianism*** and ***pragmatism*** in it owing to its emphasis on human consciousness and experience which were reflected in human actions and, which in turn, were directed towards alleviating human problems.

Though humanistic geography started on the same platform as of behaviourial geography, the two soon parted ways as, humanistic geography according to **Entrikin** concerned itself with the *'subjectivity'* of both the researcher and the reconnoitered. It digressed from the formal structures of behaviourialism which otherwise was supposed to have a strong connection with the positivist/spatial tradition and was rather considered an outgrowth of that tradition.

One of the first geographers to advocate humanistic geography was the Irish geographer, **William Kirk** as early as in **1951**. He published his ideas in his essay, ***'Historical Geography and the Concept of the Behaviourial Environment.'*** But perhaps the time was not appropriate since by then, geography was greatly inspired by the positivist tradition to initiate the quantitative revolution. Later in **1976**, it was **Yi-Fu Tuan** who argued for humanistic geography as concerned with people and their conditions. He opined that humanistic geography sought to achieve an understanding of the world through an insight into the human-nature relationship and the geographical behaviour of humankind as well as their perception about space and place. Geographical activities and phenomena were treated as the manifestation of human awareness and knowledge.

After the 1980s, humanistic geography advanced further from its early position of a critique of positivist philosophy to attack on structuralism. At the same time it developed an insightful methodology for empirical research. Two prominent streams of work were identified in humanistic geography. One stream tried to connect with the ***humanities*** by investigating knowledge that emanated from human feelings and experiences regarding being a human being on this planet. The other stream tried to connect with various philosophies of ***human and social sciences.***

APPROACHES TO HUMANISTIC GEOGRAPHY

Humanistic geography was developed as a conceptual perspective that highlighted on the thorough understanding of human-environment relationship particularly on the basis of individual or group awareness and experiences regarding different spatial units and related geographical phenomena. The main emphasis was on humans as rational being with the power to think and perceive rather than as mere responders to stimuli as was presented within the positivist and behaviourial framework. According to **Ley** and **Samuels**, humanistic geography incorporated a wide range of philosophical approaches within it ranging from idealism, existentialism and hermeneutics to phenomenology; the connection with pragmatism has already been mentioned before. At the same it ascribed a central role to human beings and was a people's geography with human development as its principal objective.

Humanistic geography imbibed in it the philosophy of ***existentialism*** that urged on human quality and subjectivity. It was based on the doctrine of *'existence before essence'* which implied that humans existed first and, thereafter were responsible for their every action. It stressed upon personal freedom, personal decision-making and personal commitment. In other words, the purpose of humanistic geography in its affinity with existentialism, was to analyse the existential space as occupied by humans and the ways they defined their relationship with their space. This approach was essentially historical in that, it attempted to reconstruct space through the experiences of its denizens.

As a counter to the postulates of positivism, **Leonard Guelke** propounded the philosophy of ***idealism*** and urged the human geographers especially the historical geographers to probe into *what* humans, as decision-makers believed in and not *why* they believed. Thus, human geographers were not supposed to engage themselves in developing theories as, the pertinent theories that resulted in the geographical activities under study were already extant in human minds. Humanistic geography inspired by the idealist philosophy upheld that reality was basically a mental construct and a pattern of human behaviour actually reflected the underlying ideas. Idealism according to Guelke was based on *two* propositions---(i) a ***metaphysical proposition*** which asserted that an idea or a mental construct had a particular duration which was however, independent of material things and processes; and, (ii) an ***epistemological proposition*** which believed that knowledge was derived indirectly from the subjective human experience of the world and was an outcome of human thoughts and ideas. It upheld that the existence of a *'real'* world was actually mind-dependent.

Idealism was basically a sort of ***hermeneutics*** that dealt with the theory of interpretation and clarification of meaning. It developed in the German tradition of *'geisteswissenchaften' or human science.* The contention between the objectivity and subjectivity of human discourses led to *'double hermeneutics.'* Hermeneutics was applied in contrast to the positivist-spatial science methods as advocated by humanistic geography through, a presuppositional approach directed by social conscience. It provided an epistemology that aided in restructuring regional geography by speaking of the spatio-temporal aspect of a region. At the same time, it expressed its concern regarding any spatial unit with respect to its culture as developed by humans occupying it over time particularly language.

In the 1970s, another philosophy that was more popular among the human geographers than idealism was ***phenomenology***. Though the term was first used by **Sauer** in the 1920s, it became widespread in the 1970s when **Relph** tried to introduce the approach within geography. The objective was similar to the above approaches---to present a critique of the positivist tradition. It presented an alternative to positivist philosophy that was based on the premise that there can be no objective world without human existence. **Kirk** in 1963 identified *two* different yet mutually dependent environments (i) a ***phenomenal environment*** that included everything on this planet; and, (ii) the ***behaviourial environment*** that was the perceived and experienced part of the former. Phenomenology in geography was concerned with the phenomenal environment the elements of which were considered distinctive for every human since, it was the outcome of individual perception and action. Therefore, the phenomenological approach in geography sought to explore how individual human being structured the environment in a subjective way.

THEMES AND METHODS OF HUMANISTIC GEOGRAPHY

Humanistic geography originated as a perspective against that form of human geography that was reduced to an abstract study of space and structures. Sometimes, humanistic geography could be used interchangeably with humanism because it accorded central role to humans. But

precisely, humanistic geography was mainly concerned with the outcomes of human activities. According to **Ley** and **Samuels**, humanistic geography was based on three basic precepts----------------(i)

anthropocentrism; (ii) subjectivity; and (iii) the concept of place.

Humanistic geography did not consider humans as mere *'economic man'* but attempted to investigate as to how geographical activities and phenomena were a manifestation of human awareness and creativity. As a propounder of humanistic geography, **Tuan** identified the following *five* **major themes** of humanistic geography:

- ***Geographical knowledge or personal geographies***: Humans were to be treated as rational beings with the ability to think and perceive. The main task of the humanistic geographers therefore, was to study the ideas and thoughts that emanated from human minds since these ideas constituted geographical knowledge. Each and every human being possessed such knowledge though their perception varied. They utilized their geographical knowledge for their biological survival. Hence, geographical knowledge was conceived as personal.

- ***Role of territory and creation of place identities***: As mentioned earlier, sense of place was an intrinsic aspect of humanistic geography. Every human being occupied and utilized some space with which they developed a strong sense of emotional bonding. Much of his biological needs were satiated in that space. Hence, a particular space constituted the territory of humans which was not only a confined area in its literal sense but a place with which human beings identified themselves. It was here where humanistic geographers stepped in to analyse how a mere spatial unit turned into a place identity for individual human being.

- ***Crowding and privacy***: Crowding of a place resulted in physical as well as psychological tensions which were eased out by cultural, social institutions and infrastructures. In a similar way, privacy and seclusion also influenced the thought processes and actions of humans. Privacy was thought to be required by every individual. Within the private space individuals developed their own personal world.

- ***Role of geographical knowledge in determining livelihood***: For sustenance humans engaged themselves in economic activities. They utilized their geographical knowledge to decide their economic activities. Thus, accordingly they planned their action for sustenance which was the essence of pragmatism. In doing so, they were in an position to distinguish between life-sustaining and life-destroying activities.

- ***The impact of religion***: Religion was supposed to be subjective and associated with the normative elements of values, beliefs or ethics. Religion was conceived as the desire for coherence. The variation in this desire, which differed with individual persons and culture, provided a field of investigation for the humanistic geographers.

Four conceptual and methodological themes were identified as inherent of humanistic geography.

- According to humanistic geographers, human life and experiences were regarded as dynamic and multivalent that had cognitive, attitudinal and emotional elements attached to them. Humanistic geographers asserted that the task of a comprehensive human geography was to identify these elements and understand how they contributed to human experiences and actions, as well as, how each of these elements were connected to each other in a supportive or contradictory manner. This may be made clear in the words of Tuan that every individual human was at the same time a biological being, a social being and an inimitable personality and all these three aspects were believed to be a function of environmental perceptions, attitudes and values.

- Since human experiences were indefinable, humanistic geographers departed from the scientific methods employed under the positivist regime in which everything was to be explicated and verified empirically using statistical techniques. On the contrary, humanistic geographers adopted the ontological-epistemological perspective to encompass a much wider range of experiences; which would have created a framework within which the investigators would be able to study the experiences of their subjects with greater precision.

- The humanistic geographers advocated that humanistic geography should have originated from the self-knowledge and first-hand experience of the investigator. At the same time, it should have also incorporated the experiences of the *'others.'* The others could range from people, places, any natural phenomena or any aspect of human-environment relationship. This approach of humanistic geography brought them in sharp contrast with the objective approach of the quantitative paradigm in which the experiences of the researchers were greatly minimized. With regard to this, Tuan asserted that through an understanding of geographical experiences individuals developed a sense of environmental humility and acted more compassionately towards other humans and the place and the environment they occupied.

- Humanistic geography employed the usage of ***two complimentary research methods*****----**

one that involved the *explanations of experiences*. It was based on a multitude of descriptive sources like first-hand experiences of individuals, archived reports and literature, evidence gathered through photography, films or any other forms of media. Its emphasis was to highlight the commonalities that existed in experiences related to a place or an environment. The other method that involved the *interpretation of the social world* was based on philosophical arguments rather than experiential evidences. It involved a wide range of philosophical traditions ranging from existentialism, pragmatism, idealism to post-structural Marxist approach.

CRITICAL APPRAISAL OF HUMANISTIC GEOGRAPHY

Human experience and human actions have always been the focus of humanistic geography. The central thesis of humanistic geography was provided by the criticisms rendered against positivism. It ardently highlighted upon human as *'living, thinking and acting being'* and insisted that human conditions could only be suggested through humanistic endeavours expressed in human attitudes, impressions and sense of place which otherwise could not be articulated through positivist methods.

However, humanistic geography was not free from criticisms.

- The first and the basic criticism rendered against the humanistic methods is that the researcher was not in a position to ascertain whether the real and the true explanations had been provided or not. It is true that humanistic explanations could not be established with certitude but this again provided a field of criticism by the positivist-quantitative approaches where everything could be verified empirically and thus had a greater certainty. In fact, the natural sciences whose methods were adopted by the positivist regime were mainly comprised of theories that were abandoned through further research which in turn enhanced the scope of study and resulted in more authentic and powerful theories. But with humanistic methods this was not possible.

- Secondly, on methodological grounds humanistic geography differentiated and distinguished between physical and human geography which diluted the core of the subject and gave rise to some sort of dualism in the discipline of geography. This dualism sometimes proved to be detrimental in the development of geography. Physical geography mainly dealt with inanimate objects and so its methods were mainly scientific and mathematically verifiable. On the contrary, since human behaviour was difficult to predict and varied over space and time, such quantitative techniques were not always applicable in human geography. However, humans as the prime focus of humanistic geography and physical environment of physical geography were not mutually exclusive but rather related and, could not be studied independent of the other.

- Humanistic geography was criticized as *'methodologically obscure'* since its main focus was on subjective rather than objective research. Humanistic geography was largely based on the experiences and perceptions of the humans which were mental constructs and were essentially abstract in nature with no practicality as such. Any method was acceptable to interpret the meanings of human experiences. Thus, humanistic geography had no sound or valid methodological base on which the theories developed by it could be successfully and authentically grounded.

- This gave rise to another criticism against humanistic geography that it had limited applied aspect. The investigator could have numerous interpretations of reality and in that situation it was really difficult to ascertain reality. Under such circumstances, it was rather challenging to identify the geographical problems and frame alleviating policies accordingly.

- Though humanistic geography attempted to combine several philosophical traditions along with an incisive methodology, yet as pointed out by **Entrikin**, it failed to provide a suitable and viable alternative to the scientific methods. It was better described as some kind of critical philosophy that originated against the positivist tradition with a purpose to revive the *'humane'* element in geographical research.

- The concept of place as enshrined in humanistic geography was static and exclusive. This was criticized by several post-structuralist geographers who presented a progressive and dynamic concept of space that was responsive to wider social and environmental contexts. The sense of place of humanistic geography was also questioned by the post- modernists on the ground that the distinction between perceived and real space was no longer valid in the world of booming hyperspace comprised of digital environments and virtual realities.

Humanistic geography has been subjected to criticisms and rejection by modern day geographical research due to its unscientific character and its associated gross inability to provide generalisations, laws and theories. However, since any philosophy is largely an outcome of thoughts originating in human minds, the importance of human ideas can in no way be undermined. It is true that post 1990s humanistic geography disappeared as a distinctive sub-branch of geography, but interests in humanistic themes still persists particularly among the phenomenological philosophers regarding the phenomena of space. Interestingly with time humanistic geography with its continued focus on human action, human beliefs and awareness; human interaction with their place in space and, the interpretation of that place within space, have adopted psychoanalytic theories. The objective behind this has been to do away with the criticisms regarding their obscured methodological and poor theoretical base. It had also started focusing on the increased interaction between human and physical geography particularly in determining the role of individuals' perception in creating the physical landscape.

RADICALISM

I- Introduction

In the Leftist ideological group, there were two sections – Left-liberal and Left-radical. Both the sections were concerned with inequalities, deprivation etc, i.e. problems pertaining to rich-poor divide. They were against capitalism. They both criticized Positivists because Positivism could not answer the questions of deprivation, dislocation, crime, problems of female issues, class differences etc.

Left-liberals were those people who want minor adjustment in society for the benefits of have-nots. But Left-radicals wanted to change the entire social order.

II- Context

Amid mass demonstrations against government's social policies, for which people came out on the streets of American cities, political radicalism through the revival of socialist parties happened due to certain reasons, as follows:

1. After the World War-II, there was a steady economic growth for two decades. Then an economic slowdown or slump started to happen. In such a situation, people became conscious of the role of the government – its successful schemes and failed projects. It was widely felt that the fruits of economic growth were not shared equally, and a substantive chunk of society was facing economic hardship. This fueled grievance against the government, and that's how Civil rights Movements took place in almost all American cities during the late sixties.

2. Another point of discontent was Vietnam War where USA's aggression was viewed as an imperialist hegemonic pursuit.By and large, it was against the essence of democracy which the USA preached and practiced. It not only led to the destruction of lives and properties of Vietnamese but also led to the death of US soldiers fighting in Vietnam. American people, especially the youth, revolted against the government for Vietnam War. Student protests were not only limited to its place of origin i.e. USA, but it expanded to several European countries also.

3. Problems of Black population, who lived in the shabby physical environment, started to emerge. Such problems pinpointed the failure of economic growth-centric government policies, which was running under the profit maximization policies.

III- Social Relevance Revolution

Given this context, a reassessment of purpose and methodologies of natural as well as social sciences began. It was felt that human being and their environment as a part of the earth is the most important subject that natural and social sciences should enquire in details. Geographers, who were working on the themes of "optimum location" of infrastructural facilities, now started to focus on the physical and social environment that surround people. This phase of revolution in geography, after the much-acclaimed *quantitative revolution,* is known as "radical revolution" or "social relevance revolution".

III A-The Radical Stream of the Relevance Movement:

In fact, "radical revolution" emerged as a critique of *quantitative revolution*. During the 1950s, the philosophy of positivism and empiricism became very influential. Geographers, while interacting with other disciplines, also started applying various tools and techniques to analyze and explain the spatial variation of man-nature interaction. They got so engrossed in model building, that the theoretical approach towards looking at socio-economic problems was sidelined, and availability of data and application of techniques started to guide research procedures. In such a context, through "radical revolution", a new discourse started that reminded geographers it is theoretical understanding that shows the path of research through an exploration of suitable dataset and methods, and not the other way round.

Radical viewpoint started through William Bunge's work who wrote about Radicalist ideas in his book *Theoretical Geography* in 1962 and who founded Society for Human Exploration at Detroit in 1968. This Society urged geographers to undertake fieldwork in areas where poorest people live or the areas which are most backward and depressed. Such expeditions targeted to acquire firsthand and unbiased information of these areas so that a collective engagement with local people can bring meaningful inputs and bring about sound policy and planning framework. Few expeditions were carried out in Detroit. For providing training to aspirants who shown interest to participate in such expeditions, a course was opened at the University of Michigan. As university officials did not cooperate at the later stages, such expeditions were stopped in the USA. However, the expeditions continued in Toronto (Canada), Sydney (Australia) and London (England). Moreover, the Union of Socialist Geographers (USG) was established in 1974. Members of USG also participated in special sessions of AAG conventions and IBG annual meetings.

Radical ideas flourished in the hands of David Harvey and Richard Peet. Harvey wrote *Social Justice and the City* where he talked about Black people living in Ghettos. Richard Peet started to publish articles in a famous journal known as *Antipode* in Clarke University in Massachusetts in 1969. The issues in *Antipode* were quite revolutionary. They talked about urban poverty, discrimination against Blacks, feminism and cruelty against women, crime, deprivation, problems pertaining to minorities etc. Therefore, geography again got a breakthrough from its original systematic or regional approach when it started incorporating new social issues.

Due to increased poverty and inequality, especially poverty among the people of Ghetto and rural areas, Radicalists tried to perceive planning from a new viewpoint i.e. planning *with* the people rather than planning *for* the people. According to Harvey, geographers should consider the question as to who is going to control whom, in whose interest the controlling is going to be exercised and if it is exercised in the interest of people, who is going to take it upon himself to define that public interest.

Important features and objectives of the radical stream of relevance movement are following:

1. To expose the issues of discrimination, deprivation, inequalities, crimes, issues pertaining to health and mental degradation in the capitalistic society.

2. To pinpoint the weakness of Positivism and Quantitative Revolution in geography which emphasized geography as a spatial science and did not deal with the human issues?

3. To remove regional inequality

4. Radicalists opposed economic and political concentration, imperialism and nationalism.

5. They opposed superiority of a particular race.

6. They also prescribed revolutionary changes in the work order to develop a tension-free peaceful environment for all.

Radicalism was developed as a critique of existing models, because such models especially those adopting a positivist methodology which was supposed to be value-neutral, was helping the imperialistic forces to maintain the status quo. Geography was a tool for imperialists. Radicalism was critical to this system.

Radicalists always talked from the standpoint of those people who were not in control of means of production (land, labour, capital, organization), and they always supported the downtrodden group of society. Radicalist thinking always went against nationalism. Before Radicalist thinking, geography was a science which protected the ideology of majority who owns the means of production. Radicalists criticized this scenario. It also developed as a protest against data. Radicalists thought of a society which is controlled by all.

Radicalists like James Blaut (1970) attempted to link the issue of *imperialism* with capitalism. Imperialism denotes domination and subordination of one country to another – be it in economic or political terms. More developed countries had a tendency to control less developed ones, by exploiting natural resources and setting terms of trades often biased against less developed countries. Capitalistic countries, through this control, created a monopoly situation. Another issue was *ethnocentrism*, where an ethnic group was considered superior to another group(s). European ethnocentrism pointed towards the superiority of Europe over Asia and Africa, the superiority of Whites over non-Whites. It also showed the world how and why development persists in Europe. Blaut was very critical to this unicentric model and explained how Europe progressed at the cost of disrupting African and Asian countries. Imperialistic hegemony, through colonization of many African and Asian countries, paved the way for European countries to access billions of wealth. This led to the progress of Europe, in terms of expansion of industries, commercial activities, education, and technology. As the entire idea of racial superiority and ethnocentrism was based on certain prejudices, Radicalists opposed it.

Radicalists also opposed the way females were oppressed in developing and developed countries. Females were found to have an unequal role in terms of decision-making in households. They were systematically exploited, as their role was defined from a male perspective. They had relatively less mobility, and their role was defined to be restricted within household – cooking and taking care of children and so on.

To a certain extent, Radicalism was linked with *anarchism*. Anarchism called for the removal of state, and its replacement by voluntary groups of individuals. These individuals could work without external pressure and maintain social order. In a way, anarchism promoted individual liberalism and socialism. Peter Kroptokin and Elisee Reclus elaborated on the way by which such social orders can be maintained. Kroptokin attacked capitalism on the ground that it increased competition and inequalities. He commented that mutual cooperation and support help a community or a society to live peacefully. Cooperation based production, decision-making at grassroots level, the spread of democracy, greater integration of short-distanced workspace and living space were some of the ideals many Radical geographers followed.

III B-The Liberal Stream of the Relevance Movement:

Liberalism, although beliefs in democratic capitalism, advocates executive actions for minimizing social and spatial inequalities in the levels of human well-being. It shows a commitment towards ensuring a basic minimum level of standard of living for all. In this context, it prescribes state action in helping less privileged section of the human society. Statistical techniques, involving multiple variables, were applied to map levels of human being (Thompson and associates, 1962). The work done by Smith (1973) and Knox (1975) are often referred in geographical literature. Measurement and mapping of variables related to human well-being became important, and such variables were categorized into three sub-sets- “physical needs” (nutrition, shelter, and health), “cultural needs” (education, leisure, recreation and security), and “higher needs” (through surplus income). Their works show that geographers can play a significant role in informing policy-makers about the spatial implications of inequalities so that better decisions can be taken for improving policies and schemes further. Another part of these works is raising awareness among citizens so that they become better informed on welfare issues. Cox (1973) and Massam (1976) looked at how efficiently public services can be provided, by redrawing of administrative boundaries or changing the location of public facilities. In his seminal book, *Human Geography: A Welfare Approach*, David Smith (1977) focused on “who gets what, where and how”, and this reoriented the goal of human geography towards making a society where spatial malpractices and injustices are done away with. Therefore the “distribution” of fruits of economic growth emerged as an important issue.

Moreover, this stream of social relevance movement advocates that our surrounding environment should be looked after well. Therefore, issues like environmental degradation-conservation-management are often discussed.

III- Towards Marxist Geography:

A great contribution of guiding geography towards Marxism happened through the works of David Harvey. In his book on Ghettos located in American cities, he pointed towards roots of problems that lie in capitalism. According to him, the capitalistic system created such a market-based mechanism, that regulates land use, and this is biased against the poor Black population. He argued that once a geographer adopts Marxist approach towards looking social problems, he or she cannot detach himself or herself. That's why a political awareness is generated within them, and they get actively involved in making a society with more justice. Harvey's influence was so strong that some practitioners of social relevance research started adopting a Marxist approach. Nowadays, radical geography is more aptly known as Marxist geography.

IV- Critique:

Social relevance movement, especially radicalism was able to usher in some fruitful changes in the methodological discourse of geographical studies. These are:

- From the rhetoric of quantitative technique based analysis of geographical attributes, it reoriented human geography towards prominent social and environmental issues, thereby broadening the scope of geography to interact with other disciplines of social sciences.

- The classical tradition of fieldwork in a small region was altered in the sense that more in-depth and participatory planning oriented studies were encouraged. This fieldwork entailed a new pattern where respondents were involved in the process of surveying. One needs to understand that this new pattern was quite challenging. The expeditions, promoted by the Society for Human Exploration, could not go on extensively due to multiple reasons (including existing power structure), even though it received a certain chunk of academic interest.

Some of the limitations or weaknesses of radicalism are:

- First criticism came from Russians who claimed themselves as true Communists and Marxists. Radicalism was entirely an American enterprise. Though Radicalists in America talked about the social change they never talked about an armed revolution which is a basic component of Marxist ideology.

- The theoretical base of radicalism was very weak. Basically, they were dependent on other social sciences. Whatever Harvey discussed in *Social Justice and the City* were basically sociological, political or economic analysis.

- Though the topic of radicalism was varied, the techniques and methodologies were not very path-breaking.

- Radicalism gave over-weight to Marxism. Geography, by virtue of its subject matter, is a spatial science. It cannot be explained totally with the help of Marxist thinking.

- The ultimate question was 'Who will guard the guardians'? Even socialist governments following models of Marx could not solve problems of the oppressed class.

- Humanistic geographers criticized radicalists because the former gave more emphasis on people, not as an ideology like the Radicalists. Humanistic geography says that geography cannot be explained through any generalized theory. It is human-specific.

- Positivists criticized radicalism because they don't take help from any empirical science.

- After the fall of USSR and East European nations in the end of 1980s, the worldwide impression was that socialism has no value and capitalism has won its final victory. Therefore, geography is essentially a locational science which is based on empirical positivist values, which is the tool of capitalists.

BEHAVIOURAL GEOGRAPHY

By the mid-1960s use of statistical techniques in research for precision has been largely accepted by geographers. The duality of systematic versus regional geography was resolved as both were now accepted as important components of the discipline though interdependent and equally useful. It was increasingly realized by the geographers that the models propounded and tested with the help of quantitative techniques, provided poor descriptions of geographic reality as well as the man-environment relationship. Consequently, progress towards the development of geographical theory was glaringly slow and its predictive powers were weak. Theories such as Central Place Theory, based on statistical and mathematical techniques, were found inadequate to explain the spatial organization of society. The economic rationality of decision-making was also criticized as it does not explain the behaviour of man. It was a psychological twist in human geography which emphasized the role of subjective and decision-making processes that mediate the association between environment and spatial behaviour of man. It can be said that the dissatisfaction with the models and theories developed by the positivists, using the statistical techniques which were based on the 'economic rationality' of man led to the development of behavioural approach in geography.

The axiom of 'economic person' who always tries to maximize his profit was challenged by Wolpert. In his paper entitled '*The Decision Process in Spatial Context'*, Wolpert (1964) compared the actual and potential labour productivity of Swedish farmers and came to a conclusion that optimal farming practices were not attainable. He concluded that the farmers were not optimizers but, satisfies. Thus human behaviour was seen to be a product of decision-making and it was a human tendency to have incomplete information, to make imperfect choices and even then be satisfied with sub-optimal options.

2. Behaviourialism in Geography

Behavioural geography banks heavily on 'behaviouralism'. Behaviouralism is an important approach which is largely inductive, aiming to build general statements out of observations of ongoing processes. The essence of behavioural approach in geography lies in the fact that the way in which people behave is mediated by their understanding of the environment in which they live or by the environment itself with which they are confronted.

In behavioural geography, an explanation for the man-environment problem is founded upon the premise that environmental cognition and behaviour are intimately related. In other words, the behavioural approach has taken the view that a deeper understanding of man-environment interaction can be achieved by looking at the various psychological processes through which man comes to know the environment in which he lives, and by examining the way in which these processes influence the nature of the resultant behaviour.

One of the most interesting and applied aspects of behavioural geography was work examining the human perception of environmental hazards. The pioneering work by Robert Kates (1962) on floodplain management is one of the bases of this approach. He states the manner in which human beings perceive the uncertainty and unpredictability of their environment play a significant role in the process of decision making. He developed a scheme that had relevance to a wide range of human behaviour. This scheme of Kates was based on four assumptions –

1. Men are rational while taking decisions.

2. Men make choices.

3. Choices are made on the basis of knowledge.

4. Information is evaluated to pre-determined criteria.

Subsequently, Kirk (1952-1963) supplied one of the first behavioural models. In his model, he asserted that in space and time the same information would have different meanings for people of different socio-economic, cultural and ethnic backgrounds living in a similar geographical environment. Each individual of a society reacts differently to a piece of information about the resource, space, and environment. This point may be explained by citing the following example.

The highly productive Indo-Gangetic plains have different meanings for different individuals belonging to a various caste, creed and religion. Jats, Gujjars, Ahirs, Sainis, Jhojas and Gadas living in the same village perceive their environment differently. A Jat farmer may like to sow sugarcane in his field, a Gada and a Jhoja may devote his land to sugarcane, wheat and rice, an Ahir ma Saini is invariably interested in intensive cultivation, especially that of vegetables. For a Saini (vegetable grower), even five acres of arable land may be a large holding, while a Jat who uses a tractor considers even 25 acres a small holding. The perceived environment of each of these farmers living in the same environment thus differs from each other both in space and time.

The aspect which was most enthusiastically adopted by geographers from behavioural analysis was the concept of *mental maps*. The paper of Peter Gould (1966) was the seminal contribution in this regard. He points out that since decisions on location were guided by the manner in which a human being perceives the environment, it becomes essential for a geographer to have a mental image of how one perceives his environment while making decisions. Therefore, *mental maps* are not just images or maps but an amalgamation of information and interpretation that a person has on a particular thing as well as how he or she perceives that place (Johnston, 1986). This was further developed by Gould (1966), Downs (1970), Downs and Stea (1973), Gould and White (1974) and Saarinen (1979) through their writings.

Gould opines that *mental maps* are not only means of examining a person's area of a spatial preference but also provides insight into the processes which led to that particular decision. He states that *mental maps* may provide a key to some of the structures, patterns and processes of man's work on the earth surface. The conceptual framework provided by Downs (1970) has been illustrated in Figure 1. This framework proposes that information from the environment (real world) is filtered as a result of personality, culture, beliefs, and cognitive variables to form an image in the mind of a man who utilizes the environment. On the basis of the image formed in the mind of the utilizer about the environment, he takes a decision and uses the resources to fulfil his basic and higher needs. Downs' framework also suggests that there exist an 'objective' and a 'behavioural' environment.

Pred (1969) presented an alternative to this inductive approach of behaviouralists on theory building on 'economic man'. In his work *Behaviour and Location,* he proposes a *behavioural matrix* (Figure 2) to give a structure in which decisions of locations can be analyzed. The axes of the matrix are quality and quantity of information available and the ability to use that information; man as an economic being is at the right-hand corner. As there is variation in the quantity and quality of the information, the position of man on the axis would also change. His position would reflect his aspiration levels, experience and even norms of the group to which he may belong. His further states, that even same individual would not be in the same position as his decisions may vary over time as spatial patterns are never static in nature.

During the 1970s, a range of related personality assessments, such as personal construct theory and the semantic differential were employed, and in this work geography and psychology became close neighbours (Aitken, 1991; Kitchin, Blades and Golledge, 1997). In particular, this productive interdisciplinary relationship was developed through the annual meetings of the Environmental Design Research Association and in the pages of the new journal, *Environment, and Behavior*. Since that period, behavioural geography has continued to diversify, even if its position has been less elevated than in the 1960s and 1970s when many disciplinary leaders worked in this sub-discipline. More recent research has included analysis of environmental learning, spatial search, developmental issues in spatial cognition and cartography and Golledge's (1993) important work with the disabled and sight-impaired. But some of the lustre has left the field. In part, this may be related to the methodological sensibilities of post-positivist human geography. In part, it is due to the growing conviction of the inherently socialized nature of geographical knowledge, which challenges the individualism of psychological models. In part, it emanates from a suspicion of the adequacy of an epistemology of observation and measurement that may leave unexamined non-observable and non-measurable contexts and ideological formations. Nonetheless, behavioural geography has a continuing legacy, comprehensively itemized and integrated into the massive compilation of Golledge and Stimson (1997).

3. The objectives of behavioural approach were:

1. To develop models for a human phenomenon which would provide an alternative to the spatial location theories developed under the influence of positivism.

2. To define the cognitive (subjective) environment that determines the decision-making process of humans;

3. To come up with psychological and social theories of human decision-making and behaviour in a spatial framework;

4. To change the emphasis from aggregate populations to the disaggregate scale of individuals and small group

5. To search for methods other than those popular during the quantitative revolution that could uncover the latent structure in data and decision-making;

6. To emphasize on procession rather than structural explanations of human activity and physical environment;

7. To generate primary data about human behaviour and not to rely heavily on the published data;

8. To adopt an interdisciplinary approach for theory-building and problem-solving.

The fundamental arguments of the behavioural geography to achieve these objectives are that:

(i) People have environmental images;

(ii) Those images can be identified accurately by researchers; and

(iii) There is a strong relationship between environmental image and actual behaviour or the decision-making process of man.

The behavioural paradigm has been shown in Figure 3. In this paradigm, man has been depicted as a thinking individual whose transactions with the environment are mediated by mental processes and cognitive representation of the external environment. In geographical circles, this concept is derived primarily from the work of Boulding (1956) who suggested that over time individuals' developmental impressions of the world (images) are formed through their everyday contacts with the environment and that these images act as the basis of their behaviour.

4. Salient Features of Behavioural Geography

The salient features of behavioural geography are discussed in the following section:

1. The behavioural geographers argued that environmental cognition (perception) upon which people act May well differ markedly from the true nature of the real environment of the real world. Space (environment) thus can be said to have a dual character:

(i) As an objective environment—the world of actuality—which may be gauged by some direct means (senses); and

(ii) As a behavioural environment—the world of the mind— which can be studied only by indirect means.No matter how partial or selective the behavioural environment may be, it is this milieu which is the basis of decision-making and action of man. By behavioural environment, it is meant: reality as is perceived by individuals. In other words, people make choices and the choices are made on the basis of knowledge. Thus, the view of behaviour was rooted in the world as perceived rather than in the world of actuality. The nature of the difference between these two environments and their implications for behaviour was neatly made by Koffka (1935-36) in an allusion to a medieval Swiss tale about a winter travel.

2. Secondly, behavioural geographers give more weight to an individual rather than to groups, or organizations or society. In other words, the focus of the study is the individual, not the group or community. They assert that research must recognize the fact that the individual shapes and response to his physical and social environment. In fact, it is necessary to recognize that the actions of each and every person have an impact on the environment, however, slight or inadvertent that impact may be. Man is a goal-directed animal who influences the environment and in turn, is influenced by it. In brief, an individual rather than a group of people or social group is more important in a man-nature relationship.

3. The behavioural approach in geography postulated a mutually interacting relationship between man and his environment, whereby man shaped the environment and was subsequently shaped by it (Gould, 1980).

4. The fourth important feature of behavioural geography is its multidisciplinary outlook. A behavioural geographer takes the help of ideas, paradigms, and theories produced by psychologists, philosophers, historians, sociologists, anthropologists, ethnologists, and planners. However, the lack of theories of its own is coming in the way of rapid development of behavioural geography. Therefore, one can say that the behavioural approach in geography is a fruitful one and it helps in establishing a scientific relationship between man and his environment. The broad scope of behavioural geography is remarkable even by the standards of human geography.

5. Criticisms:

There are, however, overall, biases in content towards urban topics and towards developed countries. One of the main weaknesses of behavioural geography is that it lacks in the synthesis of empirical findings, poor communication, inadvertent duplication, and conflicting terminology. In behavioural geography, the terminology and concepts remain loosely defined and poorly integrated, primarily owing to the lack of systematically-organized theoretical basis.

Another shortcoming of behavioural geography lies in the fact that most of its data are generated in laboratory experiments on animals and the findings are applied directly to human behaviour. Koestler (1975) pointed to the danger of this strategy, in that behaviouralism "has replaced the anthropomorphic fallacy—ascribing to animals human faculties and sentiments—with the opposite fallacy; denying man faculties not found in lower animals; it has substituted for the erstwhile anthropomorphic view of rat, a rato-morphic view of man". In short, behaviouralist theories are elegant but unhelpful when it comes to understanding the real world man-environment interaction.

Behavioural geography has too often put too much emphasis on ego-centred interpretations of the environment. Specifically, scholars are critical of two assumptions on which a great deal of behavioural research in geography is based. The first assumption is that there exist identifiable environmental images that can be accurately measured. It is not clear whether an environmental image can be extracted without distortion from the totality of mental imagery. Moreover, not enough effort has gone into checking and validating the methods by which images are elicited.

The second critical assumption is that there exists a strong relationship between revealed images or references and actual or real-world behaviour. The main objection to this assumption is that it is an unfounded assumption because extremely little research has been undertaken to examine the congruence between image and behaviour.

Another significant deficiency in behavioural geography has been the gap between theory and practice. This has been most noticeable over the question of public policy. In fact, behavioural geographers remain observers rather than participants. There is a serious lack of knowledge of planning theories and methods amongst behavioural geographers, which is an impediment to more active involvement.

It is a barrier that can be removed only by developing the requisite understanding of the planning processes; it cannot be camouflaged by noble sentiments and moral tone. For instance, it will be only rarely that a small survey carried out upon a sample of students will supply the basis for far-reaching policy recommendations, yet the final paragraphs of many such works contain this seemingly obligatory element. Despite several constraints and methodological limitations, behavioural geography is now widely accepted within the positivist orientation. It seeks to account for spatial patterns by establishing generalizations about people-environment interrelationship, which may then be used to stimulate change through environmental planning activities that modify the stimuli which affect the spatial behaviour of us and others.

The research methods of behavioural geography vary substantially but the general orientation—inductive generalization leading to planning for environmental change— remains. Eventually, it is hoped, a 'powerful new theory' will emerge. Golledge argued that substantial advances in understanding spatial behaviour have already been made by studying 'individual preferences, opinions, attitudes, cognitions, cognitive maps, perception, and so on—what he terms processes variables.

UNIT -7: WELFARE GEOGRAPHY; FEMINISM AND FEMINIST GEOGRAPHY

Introduction

Welfare geography is an approach to geography where the emphasis is on spatial inequality and territorial justice. Destined up with the rise of radical geography in the early 1970s, welfare geography stresses the need to identify and explain the existence of crime, hunger, poverty and other forms of discrimination and disadvantage. Welfare geography sought to reveal who gets what, where and how. This early work was largely descriptive and developed the abstract formulation used in welfare economics, grounding it empirically but maintaining the use of algebraic representations. It provided a basis for evaluation. Current welfare configurations, regarding who gets what, where and how, could be judged against alternatives. This preoccupation with description eventually aimed to match, and then superseded, by work on the processes through which inequality is shaped. Marxist economics replaced neo-classical economics as the basis for illustrative analysis, which takes place at two different levels. The first involves understanding how the whole social, economic and political system functions, and teasing out universal tendencies. In the case of capitalism, this level of analysis reveals that inequality is endemic. Uneven development is the spatial imprint, the geographical result of the restlessness of capitalism as a system. The second level of explanation attends to the details of particular social, economic and political systems; for example, how housing policy under capitalism advantages some people in some places and disadvantages other people in other places. The analysis of the politics behind these policies has recently been strengthened as part of renewed interest in the relationship between social justice and the state. Accompanying an attention to the restructuring of the welfare state, which characterizes much of this recent work (Peck, 2001), have been endeavored to theorizing a relational ethics of care. Illustration of feminist theory, this work seeks to uncover the social relations behind construction of care and justice. Understanding politics as an integral part in the daily deed, the emphasis is on the connections and relations rather than the difference between categories, such as private and public, state and market (Smith and Lee, 2004).

Welfare geography focuses on the connection between the spatial variation of need and structures of a provision in the creation of geographies of welfare (Smith 1973). A rather late arrival of welfare approach in humanities and social sciences and particularly in geography has several political, historical, and psychological reasons, e.g., the Vietnam War, crime explosion, environmental degradation. The manifestation of social, political and economic injustice through these crises in cities and towns led a group of social scientists to come up with a new idea and promote the radical approach. Especially, with geography, the issue of distribution was taking new urgency (Smith, 1977). Before the dawn of the Quantitative revolution, geography, like all other main sister disciplines from Humanities and arts, faced many philosophical and methodological problems. Geography did not progress as a well-regulated discipline. In the recent years, geographers have, however, adopted new strategies by restructuring their courses of the study and designed the themes around contemporary issues like socio-economic development, rural-urban studies, making the subject a primary source of awareness of local surroundings and regional milieu. During the last five decades, the subject matter of geography has experienced immense changes in the subject-matter, philosophy, and methodology. The issues of primary concern on which the geographers are concentrating nowadays include hunger, poverty, racial discrimination, pollution, environmental pollution, social inequality or injustice and use and the overuse of depleting resources, etc.

Some of the leading works and issues which have been useful in the public policy making are Black-Ghetto, Geography of Crimes and Geography of Social Well-being. The quantitative revolution of the 1960's infused a vigor into geography, which was vastly essential for the indepth and comprehensive analysis required in any public context and the formulation of proposals for public policy.

Scientific revolution paved its way in geography in the early 1960s. The pragmatists introduced the use of scientific methods (positivism) for finding solutions to the problems faced by human beings. It is with this intent that scholars like David M. Smith have embraced the welfare approach while debating the problems, prospects and the future scenario of human geography.

The welfare approach in geography has been defined differently by some eminent scholars of geography. Mishan was of the view that, "theoretical welfare geography is that branch of study which endeavors to formulate positions by which we may be able to rank, on the scale of better or worse, alternatives in the geographical situation open to society." In the spatial context, Smith defined welfare geography as the study of "who gets what, where and how."

The geographers whose prime concern are the problems of society and are trying to formulate more realistic plans for public policy by giving the description and explanation of the phenomena. Through such analysis, they evaluate their plans and suggest suitable strategies for the balanced development.

The explanation involves the empirical identification of territorial levels of human development and the human condition. This is a major and instantaneous research area in which astonishingly little work has been done in India and other developing countries as well as developed Explanation covers the how? It involves in identifying the cause and effect relationship links among the different activities undertaken in society, as they contribute in determining who gets what and where. This is where the analysis of the economic, demographic and social patterns mentioned above logically fits into the welfare structure.

Geographical distance and ease of understanding mean that some people will be enjoying the better place to for advantages or disadvantages, whether the structure is a road, railway, hospital, school, theatre, community hall, cinema, park or a recreational place. Therefore, locational decisions and comprehensive plans for spatial allocation of resources must be made with utmost care and dedication, if the benefits and penalties are found to be proportional among the population in a more or less predictable and reasonable manner. In such public policy decision making, geographers' role becomes authoritative as they have the necessary expertise in the Spatio-temporal analysis of any such phenomena.

Spatial allocation problems are related to the identification of priority areas, planning routes, the location of factories or other sources of employment, the spatial arrangement of facilities providing medical care, housing complexes, shopping centers and allocation of land for different urban and recreational uses. Each of these decisions could be made in some ways, and every decision can have a different influence. Geographers by their expertise can build up more sophisticated knowledge and models of the process of development. This involves unscrambling and complex networks of economic, social and cultural relationships and also the ecological relationships in equilibrium, so easily disturbed by ill-conceived 'developmental' projects.

Geographers by sharing out, analysis, and synthesis of space can contribute, successfully, meaningfully and more efficiently to the formation of the policies for the public, property, etc. In developing countries like China, India, and Brazil there is relatively a high degree of internal inequality. On the other hand in the Third World nations, wealth and power and other facilities of public interest are still largely in the hands of urban elites or big landlords. In India, more than 50 percent of the population is still below the poverty line and on the contrary over 50 percent of the total national assets are in the hands of only a dozen families. Moreover, in India, most of the economic activities are concentrated in metropolitan cores, although still, more than 70 percent of the total population is residing in the rural areas. The urban-based industrial and social infrastructural policy adopted by planners is widening the already wide gap on the one hand, between the rich and the poor and on the contrary between rural and urban population. The highly advanced and developed countries like U.S.A., Russia, Australia, and Japan also have spatial disparities in levels of human development. In the United States, the overall material standard of living is higher than anywhere else in the world. Millions of Americans, especially Negroes (black people), live in poverty and social denial in ghettos (city slums). In many parts of the rural south of U.S.A. (Texas, Georgia, etc.) the living conditions of some people are as bad as anywhere in the African continent. In these ghettos, the rate of crimes like drug addiction is very high.

The perseverance of widespread poverty in American slums, the most affluent society in the world is a paradox which underlines the failure of economic growth under a capitalist system to uplift the lives of all people to a current standard of decency. In 1976, according to the U.S. Census Bureau, about 12 percent (26 million) Americans were below poverty line.

One of the opinions put forward by the capitalist for the existing regional inter-regional and intra-regional disparities is that human beings are not born equal, and hence they cannot be equal in their societies. This situation gets further serious if the social, political and economic organization is intended or formed with an urban-biased or rich people-centered policy. The planners with the help of geographers can construct general social amenities which can benefit all sections of the society.

Geographers, however, cannot be the cure for all the ills, inequalities and socio-economic imbalances that are persistent. Geographers can analyze the spatial dimension of environmental problems, natural hazards and more particularly they know how to handle, analyze and interpret spatially distributed data. This consciousness of and facility of tackling the spatial dimension, which is a major component of all problems of resource and environmental management, is something not provided by those in other disciplines and have a tendency to be overlooked if a geographer does not arrange it. A welfare society needs better sharing of commodities, better distribution of commodities and better of means of manufacture among individuals (groups or classes) and places. All these things are more easily attainable if geographers who are dealing with the man-environment interface and elaborate the spatial distribution of phenomena are actively involved in the procedure of planning and formulation of public policies at different levels of development, i.e., the local, regional, national and international levels.

In countries like Sweden, the Netherlands, France, Norway, Israel, Denmark, U.S.S.R., Australia

and New Zealand where geographers in collaboration with the scholars and scientists of other fields to design public policies. Which is effective and beneficial and reaching all sections of the societies. Similarly, Geographers in India can also provide practical proposals for solving the various social, economic and infrastructural problems that are caused by rapidly increasing population.

Difference between Welfare Human Geography and Humanistic Geography

By their efforts, geographers can bring the causal relationships between inequality arising between the spatial organization of society and social structure. Public policies about restructuring and rearrangement can be designed properly by the experts through planning.

Although, human geography has appeared from earth sciences and has persistent links with physical geography. Hitherto the core aim of this particular branch of knowledge is to examine the various problems of different social groups about their environment. At present, especially after the 1960s, the geographers have adopted welfare approach as a go-to approach for the study of the human behaviour. The welfare approach, in fact, emerged as the response to quantitative revolution, spatial science, positivism, and model-building which was thought to be unsatisfactorily concerned with contemporary glitches of human societies.

The 1970s saw a chief redirection of human geography in the direction of 'welfare' issues such as hunger, deprivation, malnutrition, poverty, crime, distribution of income, assets, and access to social services (e.g., education and healthcare). This corresponded to a major change in social concern, from constrained economic conditions of development or progress to wider aspects of the quality of life. Spatial distribution of phenomena and distributional issues have presumed new importance in the present era of slow economic growth, for in these conditions policies of distribution in favor of the poor or socially deprived can be instigated only at the expense of the rich or better-off members of society. This is also known as Pareto optimality—a condition in which it is not possible to make some people rich without making others poor. The Pareto model assumes that one society has touched the edge of production possibilities, i.e., if there is no more growth; the poor cannot be made rich unless at the expense of the rich.

The basic emphasis of the welfare approach is who gets what, where and how. The ‘who’ refers to the population of the area under review (a city, region or country, or the entire world), subdivided into groups by caste, class, race, or other relevant characteristics like religion. The ‘what’ refers to the different goods and bads enjoyed or tolerated by the population, in the form of services, commodities, social relationships, environmental quality, and so on?

The ‘where’ reflects the fact that living standards differ according to the place of residence. The ‘how’ refers to the procedure whereby the observed differences arise. The initial task posed by the welfare approach is descriptive. The present state of society concerning the fact that who gets what? Where? Maybe signified by the extension of the abstract interpretations of welfare economics, and hence, the practical objective is to give these empirical substances to the people. In a spatially disaggregated society, the general level of welfare may be written as:

According to the Dictionary of Human Geography edited by R.J. Johnston, D. Gregory and David M. Smith (1994), “in a spatially disaggregated society, the general level of welfare may be

written as:

$W = f(S1 \ldots\ldots Sn)$,

Where S is the level of living or social well-being in a set of n regional subdivisions. In other words, it can be said that welfare is the function for the distribution of goods and bads among different groups of the people as defined by the area of residence. Social well-being may be defined regarding what people get, as follows:

$S=f(X1 \ldots Xm)$,

Where 'X' represents the quantity of them goods and bads consumed or experienced. Social wellbeing may also be expressed regarding the distribution within the area in question:

$S = f(U1 \ldots Uk)$,

Where 'U' is the level of well-being, ‘utility’ or satisfaction of each of the k population subgroups. In all the above expressions, the terms may be weighted differentially and joined according to any function, to denote the combination of territorial or regional levels of wellbeing, goods, and bads that maximize the objective function (W or S).”

For identifying a discrepancy in territorial distribution, development of social indicators is of great importance. Such indicators may be as follows: housing, income, education, employment, social orders or social participation, etc.

The welfare approach found Neo-classical economics as the least suitable one to explain social inequality. The Marxian economics provides a useful tool for analyzing social problems such as housing, income, education, employment, etc., because capitalism has an inherent tendency of to create disparity. The second level of explanation deals with the process of how specific elements of a socio-political and economic system operates. D.M. Smith (1977), in his Human Geography: A Welfare Approach, first suggested the approach which later amalgamated with other approaches of geography dealing with the issues of inequality. The issues dealt by welfare geography demand an interdisciplinary approach of the highest order. Moreover, in a rapidly changing era of globalization where the developing South stands deprived vis-a-vis the advanced North, there has been a transformed interest in welfare geography.

Welfare and Social wellbeing

The welfare geography approach deals with the issues related to inequality and injustice. The approach grew up as a reaction to the quantitative and model-building traditions of the 1960s. In the 1970s there was a major redirection of human geography towards social problems, viz., poverty, hunger, crime, racial discrimination, access to health, education, etc. The issues such as the distribution of the fruits of economic development received attention mainly as a result of dramatic socio-political changes in Eastern Europe and South Africa. Therefore, the basic emphasis of welfare geography is on who gets what, where and how. The 'who' suggests a population of an area under review (a city, region or nation)? The 'what' refers to various facilities and handicaps enjoyed and endured by the population in the form of services, commodities, social relationships, etc. The 'where' refers to the differing living standards in different areas? Moreover, 'how' reflects the process by which the observed differences arise. The empirical identification of inequality in territorial distribution involves developing social indicators. These may combine particular elements of social well-being in a composite manner.

Conditions that may be included are income, wealth, employment, housing, environmental quality, health, education, social order (i.e., the absence of crime, deviance and other threats to social stability and security), social participation, recreation, and leisure. Alternatively, the focus may be on individual aspects of social well-being, such as inequalities in access to health care or the differential experience of a nuisance such as noise, air pollution and so on. Descriptive research of this kind is justified because it provides information on aspects of life hitherto neglected in geography. It also provides a basis for evaluation, whereby the existing state is judged against an alternative (the past, predicted or planned) according to some criterion of welfare improvement. Thus, the impact of alternative plans for facility location or closure (e.g., hospitals, schools) could be judged by the test of which would most equally distribute the benefits (such as access to health care) among the population of various sub-divisions of the area under review. This raises the question of rules of distributive justice and the manner in which they are applied (explicitly or otherwise) in the political process. Although originally proposed as an alternative framework for human geography, the welfare approach has now been merged with other lines of inquiry within geography directed towards the fundamental problem of inequality.

Implicit in 'welfare geography' is recognition that the issues in question extend beyond the limits of a single discipline, and in fact, render disciplinary boundaries increasingly irrelevant. The welfare approach logically requires a holistic social science perspective. In order to achieve the welfare target, geographers are attacking social problems and exploring the causes of socio-economic backwardness, environmental pollution, and uneven levels of development in a given physical setting. Now, the main objective of geographical teaching and research is to train students in the analysis of phenomena, so that they can take up subsequently the problems of society as the fields of their research and investigation, thereby helping the local, state and national administration. Problems are being tackled with approaches ranging from positive to normative, from radicalism to humanism, and from idealism to realism.

In brief, geographers are increasingly concerning themselves with the problems of society, conditions of mankind, economic inequalities, social justice, and environmental pollution. For reduction of regional inequalities and for the improvement of the quality of life, the main concern of geographers is with what should be the spatial distribution of phenomena instead of with what it is. It is in this context that the spatial inequality in social amenities and living standards is investigated by geographers to trace the origin of disparity rather than to condemn injustice.

The geographers who are mainly concerned with the problems of society and trying to formulate pragmatic proposals for public policy clarify the description and explanation of the phenomena. On the basis of such analysis, they evaluate their plans and prescribe suitable strategies for balanced development. Description involves the empirical identification of territorial levels of human well-being i.e. the human condition. This is a major and immediate research area in which surprisingly little works has been done in India and in other developing countries.

FEMINIST GEOGRAPHY

Introduction

It is very important to find answers to certain queries before going into a detailed discussion about feminist geography as, the key concept of the discipline may be rooted in it. Several statistics across the globe pose certain questions before us as to why there are lesser number of females in certain parts of the globe as compared to males; why the prevalence of illiteracy is more among females than males; why females in younger age groups tend to be more unemployed than their male counterparts; or why females are most often under-represented in governments and politics. In short, whether in terms of birth, education, economy or politics, opportunities and power are unequal between the sexes. It is this *'inequality'* that forms the subject matter of what is known as ***'feminism.'*** The most important feature of feminism is that it challenges the traditional thinking by connecting issues of production with the issues of reproduction; and the personal with the political.

The feminist theory is essentially based on *three* assumptions:

- ✓ Gender is a social construct that oppresses women more than men.
- ✓ These constructs are shaped by patriarchy.
- ✓ Women's knowledge about these constructs helps in envisioning a future non-sexist egalitarian society.

Thus, two relevant concepts that need to be understood here are that of ***'gender'*** and,***'patriarchy.'***

The word gender is often used interchangeably with sex, though the two have different connotations altogether. While sex is biological, natural and remains constant over space and time; gender is a social construct that may vary with time, space and culture. Gender is a social classification of the sexes into masculine and feminine. Different masculine and feminine qualities may have their impact on the social and spatial relations between and among the sexes. When such relations are approached by geographers from within the realm of the principles and concepts of feminism, what arises may be termed as *feminist geography*. Since feminism always deals with women's position vis-à-vis men; there may be another simultaneous field of study within geography, that is, the *geography of masculinities*. Together they constitute what can be precisely called ***'gender geography.'***

The term *patriarch* originally derived from the Old Testament means the rule of the father (*pater* in Latin meaning father). However, the feminist use of the term was introduced by **Kate Miller** in her groundbreaking book, ***'Sexual Politics'*** in **1970**. The term may be well understood in the words of **Marilyn French** as the manifestation and institutionalization of male dominance over women and children in the family and the extension of this dominance in the society as a whole. The following aspects of women's lives may be under patriarchal subjugation:

- ✓ Women's productivity and labour power.
- ✓ Women's reproductive capacity and sexuality.
- ✓ Women's mobility.
- ✓ Women's access to economic resources.
- ✓ The social, cultural and political institutions.

To develop a proper understanding of the subject matter of feminist geography, it is necessary first to understand the true meaning of the feminist theory, its development through time, the different schools of thought that emanated within it and how its methods can be used in geography.

THE CONCEPT OF FEMINISM

Feminism as a concept is often misunderstood as an approach with extreme hatred for men and that a feminist is essentially a female. But in reality, there is no biological pre-requisite to be a feminist---males can also be feminists and in fact some are, just the way some women are not. The feminist theory upholds that inequality exists between the sexes. It has *four* notable features:

- ✓ It is intensely interdisciplinary in nature ranging across various disciplines.
- ✓ Certain themes are recurrent in it----reproduction, representation, sexual division of labour.
- ✓ It imbibes in it new concepts like sexism which are not only created to address the gaps in existing knowledge but also to describe forms of social discrimination.
- ✓ It draws upon women's subjective experience to enrich knowledge.

The idea of *'women'* as a distinct social group dates back to the 18th century. The first full political argument for women's rights and individual development was inspired by the French Revolution. At that time, **Mary Wollstonecraft** described in her ***'A Vindication of the Rights of Women'*** **(1792)**, the psychological and economic damage experienced by women owing to their forced dependence on men and exclusion from the public sphere. Over time, the ideology of feminism has passed through several ***waves or phases*** that resulted in the development of its different variants.

The ***first wave*** of feminism started with the liberal principles of individual rights and freedom for women. The ***liberal feminists*** contrasted the concept of servitude of women that was considered as *'natural'* and protested against all forms of subordination that reduced women to adjuncts of their husbands or fathers. The roots of this stream of feminism can be traced in 17th century British liberalism and the French Revolution. **Wollstonecraft**, a liberal feminist advocated for the protection of women under civil laws, their right to be politically represented and to be engaged in well-paid work and respected professions so as to reduce their dependence on the institution of marriage. **Harriet Taylor** argued that women should be allowed to work even after their marriage because, not only will her economic contribution to the family promote her status within it, but it would also enhance her freedom of choice. Domestic violence and the tyrannical behaviour by the husbands was a central theme of focus for **John Stuart Mill.**

By the 1960s, though the first wave of liberal feminism had achieved its basic goals in Europe, women still suffered from various forms of legal discrimination and were grossly unequal in both economic and political terms. The ***second wave*** of feminism thus, that started in Europe towards the end of the 1960s, sought to adopt a socialist and radical standpoint. Since 1970s, many feminists had started questioning the relevance of liberalism as a possible remedy to women's subjugation. Hence, ***Marxist feminism*** emerged as a dominant strand of feminist ideology in the 1970s and 1980s. This variant of feminism, as the name suggests, drew its ideas from the theories of **Karl Marx**. It attempted to link the situation of women's oppression to class struggle and economic development. Though Marx himself did not have much to say regarding the situation of women, his methods and concepts were universally accepted and applied. This method argued that the key to comprehend the women's question is laid in the development of production, that is, economy and technology. Therefore, like any other social organization, the relationship between the sexes is a function of a particular stage of economic development and cannot be altered on its own but only through socio-economic changes resulting from class conflict and revolution. **Engels** believed that women's oppression did not exist through time but only started with the creation of private property and a class-based society. Hence, only with the overthrow of capitalism, such oppression would disappear as, women would be no longer economically dependent on men and socialization of housework and childcare would free them from domestic chores. Therefore, women instead of fighting for their own causes should stand by working men for a revolutionary transformation of the society. This strand of feminism ruled out the idea that the interests of working men and women might conflict and that, women can have group interests beyond class lines or gender relations.

By this time, another group of feminists were developing their theories asserting that patriarchy, and not class was the oldest form of oppression. They constituted the ***radical feminists*** who originally worked within the Marxist set up in which they found that women's issues were treated as trivial. They were of the view that Marxism and feminism were not compatible with each other. However, in response to this there were some Marxist feminists who rejected the concept of patriarchy as historical and opined that women's issues could not be isolated from a wider socialist movement. They tried to analyse women's work both in home and in paid employment which eventually gave rise to the domestic labour debate and there was a demand of *'wages for housework.'* Some of the key ideas associated with radical feminism may be listed as (i) unity of theory and practice; (ii) linking the personal with the political; and (iii) the fundamental nature of women's oppression and subordination.

By the 1990s, there was a deep distrust for any metanarratives or any universal philosophy as Marx's. This was the beginning of the post-modern era. **Jean Francois Lyotard's *The Post Modern Condition* (1984)**, laid the foundation for post-modern feminism which believed that, women like race, class or ethnicity could not be used cross-culturally to describe the practices of human societies and that it was not a universal category. Lyotard criticized the Marxist philosophy for propounding a homogenous society which was believed to be created only through coercion. Post-modern feminism upheld that social identities were heterogeneous and complex, and it was thus impossible to create a totalizing social theory.

EVOLUTION OF FEMINIST GEOGRAPHY

By the 1970s it was increasingly felt that very little attention was being paid to the matter that whether the methods of mainstream research and theoretical approaches could be applied in feminist studies. Prior to this, it was a widely held notion that women were not capable of political thinking or economic decision-making and, even in academia the discipline of geography was no exception to this. It was realized that since there were very less women academicians in geography, women's issues were not sufficiently studied in it. The preliminary objective was therefore, to make women visible in the field of geographical studies. What followed was a series of articles that attempted to probe the position of and acknowledge the presence of women within geography. One of the pioneering works was ***'The Strange Case of the Missing Female Geographer'* (1973)** by **Wilbur Zelinsky**.

Drawing inspiration from the development of feminist theory in the social sciences and the welfare, radical and Marxist streams of geography, soon there were works produced by members of several women study groups and professional geographical associations in United States, Canada and Britain. Mention in this regard, may be made of ***The Women and the Geography Study Group*** of the **Institute of British Geographers (IBG)** who presented a series of researches on feminism and geography at the annual meeting of the IBG in 1981. In 1983, they also organized a series of sessions on feminism as a mode of geographical thought and thereafter in 1984, published their landmark work, ***'Geography and Gender: An Introduction to Feminist Geography.*'** In 1982, **Janice Monk** and **Susan Hanson** collaborated to produce an outstanding article ***'On Not Excluding Half of the Human in Human Geography.'*** **Mazey** and **Lee's *'Her Space, Her Place* (1983)** provided one of the best introduction to this emerging branch of geography. Taking recourse to conventional geographical methods, they tried to map the geography of women's rights; status of abortion laws; economic and political participation of women; their differential access to education, income and health services; their daily travel patters as well as long-term migration patterns. In 1984, two important works of feminist geography came forth in United States---a Ph.D. thesis was written in the department of geography at the University of California in Berkeley that was devoted entirely on feminist geography; and, a special edition of the geography journal *Antipode* (mouthpiece of the radical geographers) was published dealing exclusively with feminist geography. Following their British and American counterparts, a new specialized study group named the ***Canadian Women and Geography (CWAG)*** was created within the **Canadian Association of Geographers** in 1985. All these, greatly inspired the initiation of a multitude of research on women's topics by feminist academics in geography like—urban environment, housing, transportation, women in labour force, access to social services, violence, family structure etc.

By the 1980s, more advanced and theoretically sophisticated works began to be produced in this field. The celebrated article---***'A Woman's Place?'***—by **Doreen Massey** and **Linda McDowell** may be cited as an example. **McDowell** also published another work titled ***'Coming in From the Dark: Feminist Research in Geography'*** which itself is explanatory about the position of feminist studies in the recent past. Gradually geographical studies were being discussed more and more in feminist contexts. By this time, feminist geography was quite well- established and some feminist geographers wanted to extend the arena of this discipline beyond the Anglo-American circuit to the developing world as, in Africa, Asia and Latin America.

As the 1990s approached, feminism in geography was strongly grounded. This fact can be substantiated by the launching of a new and exclusive journal on feminism ***Gender, Place and Culture: A Journal of Feminist Geography*** in **1994**. This journal was totally devoted to issues of feminism, gender, sexuality and so on within geography.

Three interrelated observations stimulated the growth of feminist geography. Firstly was the presumption that the spatial layout is essentially gendered. To elaborate, *'private', 'home', 'suburbs'* are always associated with women in the public-private, work-home or city-suburbs relations. Secondly, it was observed that culturally specific notions about gender behaviour are greatly shaped by spatial relations. Women's access to social services is largely determined by her location and associated gender roles. Thirdly, it was found that a person's relationship to the environment is largely a function of gender. For example, the idea of safe and unsafe environment may be different for women and men.

ECOFEMINISM—A MANIFESTATION OF FEMINIST THEORY IN GEOGRAPHY

Man-environment relationship has always been one of the prime themes of geography. Ecofeminism may be viewed as a feminist perspective to the relationship between nature and humans or in short, environmentalism. During the time when feminist issues began to appear in the discipline of geography, '*green politics'* in the West also assumed the character of mainstream politics with heightened concern for an ecologically balanced earth. Both the movements---environmentalism and feminism found a common ground of subordination by *'man'* (humans in case of environmentalism and man in case of feminism)---and joined hands to give birth to a new socio-political philosophy called ***'ecofeminism.'*** The original expression of the term was '***ecological feminiane.'*** It was coined by the French feminist **Francois d' Eaubonne** to express a strong parallel between the subjugation of women in family and the society as a whole and, the degradation of nature. The term ecofeminism appeared for the first time in her book titled ***'Feminism or Death'*** in 1974. However, the term was popularized following the first ecofeminist conference that was held at Amherst in 1980 when large number of women across USA came together to launch their protests against environmental destruction. The basic essence of this concept is that, the devalued status of women in society and the degradation of nature are the two sides of the same coin. The nature was epitomized as feminine and male ownership of land and other natural resources were considered to give rise to a dominator culture. Hence, they used such terms as *'rape the land', 'tame nature'* and like. As described by **Warren (1987),** the ***basic tenets*** of ecofeminism may be elucidated as:

- ✓ Women are akin to nature whereas men are closer to culture.
- ✓ Both women and nature are conceived as *'producers of life'* that is ideologically rooted in their reproductive powers.

- ✓ A strong parallel exists between the oppression and domination of women and the degradation and exploitation of nature.
- ✓ Understanding this connection between women and nature is the basic requirement to understand the oppression of women and exploitation of nature.
- ✓ Hence, feminist theory and practice should have an ecological association.
- ✓ Likewise, environmental issues in turn, should be approached with a feminist perspective.
- ✓ Because, of the close link between women and nature, women can be perceived as important stakeholders in environmental protection and conservation.
- ✓ Finally, there should be the establishment of an egalitarian society in which there is no dominance on women or nature by *'man.'*

In the 1990s, by the time feminist geography was well-established, two prominent works on ecofeminism was produced. The one by **Irene Diamond** and **Gloria Orenstein** in 1990 named ***'Reweaving the World: The Emergence of Ecofeminism'*** laid out *three* strands of this concept:

- ✓ Social justice has to be achieved in collaboration with the well-being of the Earth as human life is dependent on the planet.
- ✓ The spiritual aspect emphasized on the sacrosanct earth.
- ✓ The third strand highlighted on the necessity of sustainability.

In the co-authored book ***'Ecofeminism'*** by **Vandana Shiva** and **Maria Mies** in 1993, they spoke of *three* kinds of domination prevalent in the world:

- ✓ Nature by humans.
- ✓ Women by men.
- ✓ Global South (the developing nations) by the Global North (the developed block)especially in terms of access to natural resources and controlling the world economy.

Shiva asserted that one of main motto of ecofeminism was to modify the outlook of the society regarding the activities and productivities of women and nature, both of whom are misconceived as '*passive'* resulting in their exploitation. **Mies** described women's work as producing sustenance and advocated that women and nature worked as partners to give rise to a new kind of relationship in which there is an essence of reciprocity. Although women usurp nature, but there is no sense of domination. Rather, there is a sense of *'to let grow and to make grow.'*

Sometimes ecofeminism is also linked with ***deep ecology*** in their contrast to male chauvinism and both being forms of radical environmentalism. The term ***'deep ecology'*** was introduced around the same time as ecofeminism by the Norwegian philosopher **Arne Naess** in an article titled ***'The Shallow and the Deep, Long-Range Ecology Movement.'*** What was common between the two were---both were critical of any kind of hierarchy, be it nature- humans, man-women, nature-culture and so forth; and, both sought to establish an egalitarian system free from any form of domination with equal right to live and flourish. But there was a basic difference between the two. While deep ecology was against *anthropocentrism,* ecofeminism was against *androcentrism.* Deep ecologists considered human population as the root cause of biospherical destruction with humans selfishly multiplying at the cost of other forms of life causing their numbers to cross the carrying capacity of the Earth. However, they failed to provide an answer to the question as to why humans reproduce even in areas with huge population size that may often lead to food shortages, overcrowding, poor health and hygiene, degradation of land, destruction of species etc. For this, the ecofeminists found their answers in a multitude of human social factors many of which were akin to issues of gender and oppression. They may be highlighted as:

- ***Sexism:*** This means the *sexual disempowerment* of women. Glorification of male virility which is an expression of sexism associates reproductive capacities and abundance of offsprings with male prestige.

- ***Motherhood:*** Social stigma is attached to women not bearing children. Motherhood is an integral part of female identity which is often considered as the most meaningful purpose of their lives. ***Cultural Factors:*** Certain forms of birth control may be forbidden and may be treated as signs of collusion by some culture. Human reproductive capacities and having children are highly valued as per some cultural norms and beliefs which again are, formulated by men.

- ***Racism and Class Oppression:*** The reproductive issues become even more complex for women of colour, poor women or those from the developing block. Imperialism has left behind a disastrous level of poverty in which the need for children is intensified by economic stress and more children are regarded as constituting a large familial workforce.

Like feminism in which lies its root, ecofeminism too is not a single line of thought. **Carolyn Merchant** described about the variants of this ecosophy (ecological philosophy) in her book, ***'Radical Ecology: The Search for a Liveable World'*** **(2005)**.

- ***Liberal Ecofeminism:*** This school of thought was in belief that if women were given equal educational opportunities as men, they could prove themselves as important stakeholders in natural conservation and improvement of the environment to ensure a higher quality of human life. They have to transcend the social stigma and biological constraints to join hands of men as scientists, lawyers, regulators and legislators for environmental conservation. This school of thought was thus in tune with reform environmentalism that sought to alter human-nature relationship within the existing framework of governance through the passage of new laws and regulations.

- ***Cultural Ecofeminism:*** Cultural ecofeminism spoke of an era prior to the emergence of the patriarchal system when femininity was held in high esteem and nature was portrayed through female deities. But, with the advent of the Industrial Revolution, the concept of the nurturing Earth was replaced with the metaphor of a machine to be controlled from outside by humans.

- ***Social Ecofeminism:*** This strand of ecofeminism advocated women's liberation through the overturning of the existing socio-economic hierarchies that translates all aspects of life to a market system. This ideology is based on the concept of *'social ecology'* that views the domination of nature by humans as an outcome of the domination of humans by humans. It envisioned of a society constituted of decentralized units that would transcend the public-private dichotomy characteristic of the capitalist system.

- ***Socialist Ecofeminism:*** This variant of ecofeminism blurs into the previous strand and, provides a critique of the dialectical relations between production and reproduction, and between production and ecology that is typical of capitalist patriarchy.

Ecofeminism has been criticized for being too idealistic in its standpoint. It overemphasized on the mystical connection between nature and women rather than highlighting the actual conditions of women. The entire onus of environmental conservation rested with women and completely undermined the role of men in maintaining environmental sustainability. Moreover, it also failed to differentiate women across space with different social background. Nevertheless, it provided a sort of platform for the achievement of *sustainable development.*

Feminist geography essentially employs the feminist philosophy in addressing several issues of social and spatial relations and the role of men and women in such social and spatial frameworks. There is a political element inherent in it that adds some sort of *'radicalism'* to this field of geography. The basic premise of *'feminist geography'* is that advocates that inequality exists between the sexes over space which should be done away with to establish an egalitarian society.

UNIT - 8: POSTMODERNISM AND POSTMODERN GEOGRAPHY

It is currently fashionable to talk of modernism and post-modernism in social discourse though precise meanings of the two terms are not always clear; so that one commentator described them as "the buzzwords of the 1980s". Important issues are, however, at stake so that the debate on postmodernism cannot be dismissed simply as "the product of disciplinary politicking, in which ascendant academics are once again jostling for new stalls in the intellectual marketplace", or as "little more than the cultural logic of late capitalism, the designer label on the baggage of flexible accumulation". But,

if these twin objections are superimposed, a different reading becomes possible and post-modernism can be seen to mark an unprecedented crisis of intellectual activity within the contemporary crisis of modernity. The debate then has the liveliest of implications for the project of a critical human geography (and critical theory more generally) because it raises acute questions about the very possibility of critique itself (Gregory, 1989, p. 348).

The two terms are closely interrelated so that it is not possible to understand post-modernism without due reference to the history and nature of modernity. Following the *Dictionary of Human Geography* (1994), *modernity* may be described as "A particular constellation of power, knowledge and social practices which first emerged in Europe in the sixteenth and seventeenth centuries, and the forms and structures of which changed over time and extended themselves over space until, by the middle of the seventeenth century, they constituted the dominant social order on the planet". Toward the end of the eighteenth century "the idea of being modem came to be associated not only with novelty but, more particularly, with looking forward rather than backward: with the so-called 'Enlightenment project' of reason, rationality and progress towards truth, beauty and just life". In the middle of the nineteenth century, Marx and Engels found an essential connection between modernity and the revolutionary dynamics of capitalism under the impact of which "All fixed, fast-frozen relations, with their train of ancient and venerable prejudices, are swept away, all new-formed ones become antiquated before they can ossify. All that is solid melts into air", wrote Marx and Engels. The beginnings of the modern world system are traced back to the sixteenth century, beginning in Europe, and organized around market exchange at international scale. But the process was greatly intensified through the nineteenth and the twentieth centuries through global scale of economic operations, so that modernity became identified as the system of market-based political-economic structure and organizations. Thus, by the second half of the twentieth century, a profound crisis of identity and representation had begun to be felt all around-in the arts, humanities, and the social sciences-particularly in the Western metropolises and the colonial capitals, owing to the accelerated pace of time-space compression. By the end of the 1950s, the modern way of viewing phenomena in ter11,s of Cartesian coordinates had begun to be questioned, separation between place and space began to be drawn, and the idea of bounded cultures through which one could journey out and analyse other cultures appeared no longer relevant (Clifford, 1988, p. 2). All this called for new approaches to explanation of social reality-approaches focused on space specificity, that is, approaches beyond the modernist tradition of universalism. This was the beginning of post-modernism in social thought. There are contrasting perspectives on the plus and minus sides of modernity, as defined above. One set of commentators holds that:

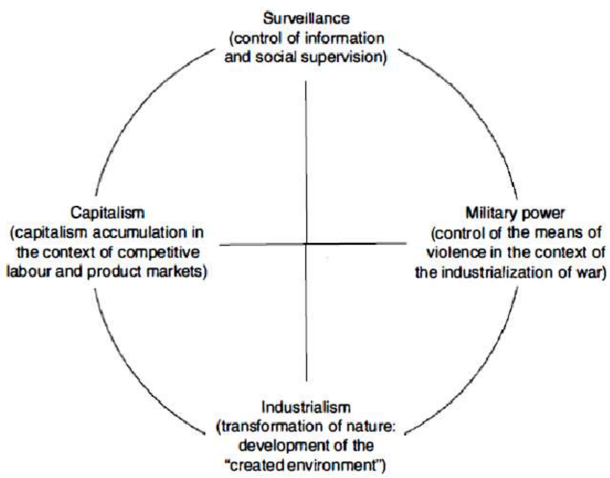
... modernity was an unqualified human good which promised to banish ignorance, misery and despotism; to free human beings from myth and superstition, disease and hunger, oppression and arbitrary rule. In the middle years of the twentieth century these assumptions culminated in models of modernization, and development programmes of which sought to remake the so-called traditional world in the image of the West (Gregory, 1994a).

But as the critics have noted, these blessings were a mixed lot. Modernism was closely intertwined with the processes and objectives of colonialism and repressive imperialism, besides ethnocentrism rooted in European cultural values and perspectives:

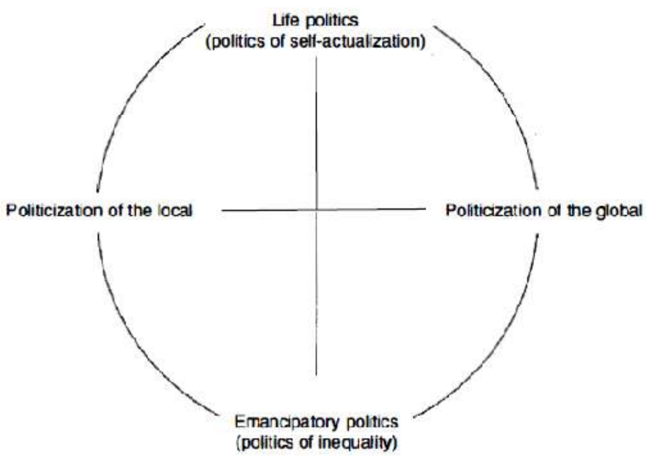
To other commentators, therefore, modernity has always had its dark side, many of them argued (in different ways) that European modernity installed new grids of power and surveillance which, in Max

Weber's well-known image, confined human agency, consciousness and creativity within an "iron cage" of· bureaucracy and regulation (Gregory, *op.* cit.).

Giddens (1990) has drawn attention to four major institutional dimensions of modernity: surveillance, industrialism, capitalism, and military power. He has also proposed an equivalent nexus of political strategies wherein the intertwining of the local and the global scales and of life politics and emancipatory politics are underlined. Figure 14.1 illustrates the proposed relationships.



(a) The institutionalized politics of modernity (Giddens, 1990)



(b) The politics of modernity

Figure 14.1: Politics and modernity (After Giddens, 1990)

Modernity, Modernization, and Modernism: The Social Science Context

In a social, scientific context Berman (1982) refers to *modernity* as "a mode of vital experience", a collective sharing of a particularized sense of "self and others", and of "life's possibilities and perils". "Modernity is thus comprised of both context and conjecture, the *specificity* of being alive in the world at a particular time and place, a vital sense of what is *contemporary.* As such, it becomes a useful general term to capture the specific and changing meaning of the three most basic and formative dimensions of human existence: space, time, and being; the spatial, temporal, and social orders of human life" (Soja, 1989a, pp. 320-321). The concept of modernity is intimately related to the multiple reconfigurations of social life that have punctuated the historical geography of capitalism since the sixteenth century. In this context, *modernization* refers to "the many different processes of structural change associated with the ability of capitalism to develop and survive, to reproduce successfully its fundamental social relations of production and distinctive divisions of labour". As Soja *(op. cit.)* points out, "these restructuring processes are continuous, but they become especially critical and accelerated during periods of deep and systematic global crisis" like the one we have entered since around late 1960s. Such periodically intensified modernizing episodes of capitalism are, in the words of Berman (1982, p. 16), shaped by:

... the industrialization of production, which transforms scientific knowledge into technology, creates new environments and destroys old ones, speeds up the whole tempo of life, generates new forms of corporate power and class struggle; immense demographic upheaval, severing millions of people from their ancestral habitats, hurtling them halfway across the world into new lives; rapid and often cataclysmic urban growth; systems of mass communications, dynamic in their development, enveloping and binding together the most diverse people and societies; increasingly powerful national states, bureaucratically structured and operated, constantly striving to expand their powers; mass movements of people and peoples challenging their political and economic rulers, striving to gain control over their lives; finally, bearing and driving all these people and institutions along, an ever expanding, drastically fluctuating capitalist world market.

In broad terms, *modernism* "refers to the cultural, ideological, reflective and ... theory-forming reactions to modernization and restructuring .... Modernism is thus the explicitly evaluative, culture-shaping and situated consciousness of modernity and is itself roughly able to be split into periods in conjunction with the historical rhythms of intensified capitalist crisis, restructuring and modernization" (Soja, 1989a, p. 322, *also see* Soja, 1989b).

The Rise of Modern Human Geography

The discipline of geography as practised over the past two centuries is essentially an "European science" which, according to Stoddart (1986), dates back to 1796, the year in which Captain James Cook first entered the Pacific. Cook's was pre-eminently a scientific voyage carrying on board illustrators, collectors and scientists whose work displayed, according to Stoddart, three features of decisive significance for the transformation of geography into an "objective science" focused on concern for realism in description; systematic classification in collection of data; and the comparative method of explanation. Stoddart maintains that it was the extension of the scientific methods of observation, classification and comparison to the domain of peoples and societies that made geography, as practised for the past two hundred years, possible. (It may be recalled here that it was George Forstera member of the scientific team that had accompanied Captain Cook-to whom Humboldt owed his interest in the study of geography.) Following Foucault (1979, 1980), Gregory (1989) has underlined that this was far from a simple "extension" of scientific principles because these

scientific principles had the most radical consequences for the constitution of the human sciences and their conception of the human subject. Besides the logic of this extension to peoples and societies "was powered by more than the ship of reason since these nominally scientific advances were at the same time the spearheads of colonialism and conquest: of subjection", so that "its trajectory cannot be separated from what Wolf (1982) once called 'the people without history'. It follows that some of the most seminal cross-fertilizations during this period were those between anthropology and geography, loosely defined, and what later became anthropogeography has to be acknowledged as a tradition of basic importance to the modern discipline". This thread of relationship is indelibly present in the writings of Humboldt, Ritter, and of course, Ratzel and Sauer. Towards the end of the nineteenth century, a second thread of relationship between sociology and geography had emerged: "the interactions between the two disciplines", Gregory comments, "were, in part, a way of clarifying their different destinations". While sociology, according to Durkheim, was to subsume the study of the spatial structure of *society-morphology sociale-the* Vidalian school insisted on the independent course of human geography predicated, on the one hand, on horizontal (spatial) relations and, on the other, on vertical relations between society and nature (Berdoulay, 1978). Around the same time, Richard Park at Chicago was airing similar views regarding sociology subsuming the study of *morphology* *sociale,* and he described geography as a *concrete science* wherein interest in spatial relations was restricted to the study of the individual and the irregular (unique), that is, to the idiographic, and not concerned with the elucidation of the general processes and universal principles, as in the case of history. Against the "concrete science" nature of geography, sociology and human ecology were described as *abstract sciences* which, even though their foundations were laid on history and geography, were essentially theory-seeking in their orientation. Thus the encounter between the two disciplines had ended in geography and sociology ultimately developing in isolation from each other (Further details on pp. 177-178). By mid-twentieth century, geography had developed close links with economics alongside the existing frontiers with anthropology and sociology. Until the late 1940s, like the rest of human geography, the human geography economics frontier was pursued in a resolutely empirical manner; but the emergence of the "new" geography in the 1950s and the 1960s gave it a pronouncedly theoretical orientation so that human geography now assumed the characteristics of the abstract sciences, in the sense that like sociology and economics, it also became a nomothetic science but (in general) it continued to eschew the study of the social element. Thus, few of the contributions in Chorley and Haggett's *Models in Geography* (1967) were addressed to social and cultural questions in any manner, and "the geometric order which they supposed to be immanent in the human landscape was derived, for the most part, from the economy: from trans codings of an abstract calculus of profit and loss into an equally abstract 'friction of distance'" (Gregory, 1989), so that human geography was transformed into spatial science. This short resume of the historiography of geography, according to Gregory, underlines three distinctive features of modern geography. These are:

1. All the three strands were bound into a naturalism-quite naturally so for a discipline that combines the study of both the human and the physical, but the other branches of the social sciences did not escape the impact of evolutionary biological thought either. "It is scarcely surprising, in consequence, that identical links should have been forged in human geography; from Vidal de la Blache's seemingly simple organicism to Hagerstrand's vision of time-geography as a 'situational ecology'; from essentially Newtonian models of spatial interaction to technically more sophisticated models of dissipative structures within abstract landscapes".

2. All the three strands were woven into a conception of science as *totalization.* Modem science holds that explanation resides in the disclosure of a systematic order whose logic imposes coherence on the apparent chaos observed by the investigator. In the study of man, this scientific tenet involved identifying a *centre* around which social life revolves. The concepts of "social system" or social structure, on the one hand, and the alternative perspective around human subjectivity on the other, served this purpose. As Gregory wrote, both these can be read in human geography, but the formalization of geography as spatial science in the 1960s accorded central place to geometry, i.e., to the search for abstract order in mathematical spaces: "Spatial science was part of an economy of explanation which put a high premium on the *simplicity* of its explanation", so that whatever be the subject of investigation, the same explanatory mould was applied.

3. If the three strands in geography since the beginning of the nineteenth century to mid-twentieth century are joined together, then human geography as spatial science would appear as a natural growth, rather than as representing any break from the past (as it was projected at the time). Geography as spatial science was "a continuation and *culmination* of modem geography", in the sense that it helped complete the search for scientific identity of the discipline as a nomothetic science. (Figure 14.2 depicts the sequential construction of human geography as a *modern* discipline.

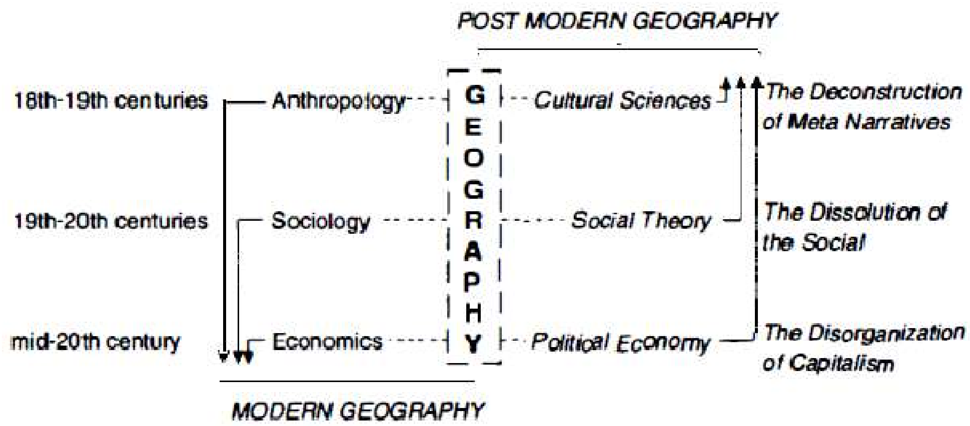


Fig. 14.2 The making and unmaking of modern human geography (*Source:* Gregory, 1989, p. 350).

THE CHANGEOVER TO POST-MODERN GEOGRAPHY

Post-modernism is a recent movement in philosophy, the arts and social sciences. Its distinguishing characteristics (according to Ley, 1994) are: Scepticism towards the grand claims and grand theory of the modem era, and their claim to intellectual superiority. Post-modernism (as contrasted to modernism) stresses openness to a range of opinions in social enquiry, artistic experimentation and political empowerr11ent. Dear (1986) classified modernism into three components: style, method, and epoch. ***Style*** is intimately implicated in the constitution of meaning and identity of the phenomenon being represented. Styles may, according to Hebd.ige (1979), either be supportive of dominant ideologies, or they may offer a ritual resistance to movements which are socially and politically opposed to the dominant view. Ley' and Mills (1992) note that the same is true about the socially constructed environment whose variations over the earth surface are of special concern to human geography. The post-modern ***methodology*** is based essentially on the strategy of deconstruction. Deconstruction is defined as a mode of critical interpretation which seeks to demonstrate how the multiple positioning of an observer (an author or a reader) in terms of culture, class and gender etc. has influenced his observation (his writing or reading) of the ***text*** (the thing being observed or read or written about). Thus, "Deconstruction is essentially a destabilizing method, throwing into doubt the authoritative claims of preceding traditions, and seeking to prise loose alternative readings of texts" (Ley, 1994). Also, post-modernism may be viewed as an ***epoch,*** that is, a historic era in which changes in culture and philosophy are sought to be located in the very process of the evolution of economy and politics at the global level. Thus some commentators regard post-modernism as the culture of late capitalism Uameson, 1984), an integral part of the new phase of post-Fordist flexible accumulation (Harvey, 1989). This, also, according to Soja (1989), is part of the attack on the predominance of ***historicism*** in modem system of thought, which highlighted the importance of individual and collective biography of diverse communities in the explanation of social relations but neglected the role of spatiality, so that the role of geographical imagination was marginalized. This subordination of space to time obscured the role of geographical interpretation in the changeability of the social world. As a consequence, geography failed to attain its rightful position in the development of theory in the social sciences (see: pp. 176-179). Historical materialistic interpretation assumed that society everywhere (irrespective of its spatial context) was characterized by similar historical pulsations in social change. Thus the social theories of historical materialism were deficient and partial in that they ignored the basic fact that the spatial context of history and culture makes a basic difference in how societies interpret particular episodes in history, and develop ways and means of adjusting to them. Consequently, the historical flow of social change was not the same everywhere. Post-modernism represented a response to modernism as a homogenizing force and to its totalizing theoreticism. Instead, postmodernism lays stress on discontinuities and disjunctures characteristic of everyday life in the real world. The homogeneity in the built-up landscape of the era of "organized capitalism" -or the Fordist era, as it is sometimes called-is contrasted with the heterogeneity of economic, social and political life in the current phase of "disorganized capitalism" (or of flexible accumulation of the post-Fordist era) (Harvey and Scott, 1989).

The post-modernist emphasis on heterogeneity, particularity and uniqueness of phenomena in differing contexts of time and space, would appear to hark back to the regionalist tradition in geography of the Vidalian era: Post-modern thought provides a theoretical context for the study of spatial/ areal diversity in the life-world. Viewed thus, post-modernism in geography is closely linked to the rise of a "new" regional geography referred to previously (pp. 256-259). The difference between the old style regional differentiation in geography of the Hettner-Hartshome era, and the present-day post-modernist regional geography lies in that, while the former was indifferent to everyday experience of societal relationships, there is now a declared commitment to the understanding of the condition of man in particular places, and the ways that spaces are socially constructed. Post-modernism is also a pronouncedly anti-spatial science in terms of the spatial scientist's search for pattern laws of behaviour. Post-modernism does not aspire to generate any grand theory of universal application; its essentially heterogeneous ethos runs counter to the emphasis on consensus on theoretical perspectives (that characterized the discourse in the spatial science era). Instead of universal theory, the emphasis now is on what Dear (1988) called, the contemporaneity of social process over time and space, so that geography, like history, is preeminently a contextual discipline, engaged in the time-space reconstruction of social life.

The movement toward post-modernism in geography (i.e., unmaking of modern human geography, Fig. 14.2) started in the 1970s with the introduction of the political economy perspective in human geography perspective that reoriented human geography to the condition of man in society and gave rise to the "radical" or "relevance" revolution. In the words of Gregory, the cutting edge of the political economy perspective was the *logic of capital* scrawling its signs on the surface of the landscape-eas against the *logic of space* (i.e., geometry) that dominated spatial science geography until the 1960s. Thus, the focus of post-modem geography became a focus on uneven development in the spatial context. There was an inevitable confusion in conceptual design in the beginning owing to the fact that political economy derived its perspective from predominantly Marxist thought, and Marx's writing, howsoever liberally interpreted, contained little about space. The rise of the political economy perspective in geography almost coincided with the emergence of what Harvey (1987) called "disorganized capitalism". Commentators now began drawing attention to the fact that the Marxist view of capitalism was deficient in that production of space is focal to the development of capitalism in the twentieth century, and the new regime of "flexible accumulation" depends upon the production of "a tense and turbulent landscape" with disjointed space economies. Such developments in space economy are fragments of a still larger mosaic:

. .. of Castell's "new relationship" between space and society, formed by the hypermobility of capital and information cascading through a vast world system, where "space is dissolved into flows", "cities become shadows" and places are emptied of their local meanings; of Jameson's "post-modem hyperspace", part of the cultural logic of late capitalism, whose putative "abolition of distance" renders us all but incapable of comprehending--of rnapping-the decentred communication networks whose global webs enmesh our daily lives (Castells, 1983, p. 314; Jameson, 1984) (Gregory, 1989, p. 353).

Towards the end of the 1970s, working out the implications of historical materialism thus moved beyond Marx's original thesis to fashion a p o s t Marxist theory of socio-spatial structuration which emphasizes that *people not only make history but also geography-which* is to say that time-space relations are intrinsic to the constitution of societies (Gregory, 1978). Such a view cuts into the sociological concept of *societies* as totalities with clearcut boundaries, and renders them extremely unstable, since under the new concept, society becomes the product of the varying spans of time and space.

This ends the earlier view of dualism between society and space, between agency and structure. Such a (supposedly) Kantian view of division between the social and the spatial was made all the more distinct in the spatial science episode in human geography. This duality (born essentially out of the urge to apply the methodology of the natural sciences to the study of sociological phenomena) had long been attacked by scholars of hermeneutics (the study and interpretation of meaning) from the time of F. Schleiermacher (1768-1834) and Wilhelm Dilthey (1833-1911). Dilthey had argued that the human sciences, owing to the peculiarity of their subject matter, required a special methodology very different from the empirical methodology of the natural sciences. This special methodology was hermeneutics that tried to search for meaning in all kinds of activities and objects, including tools and landscapes as well as individual biographies. Hermeneutics went beyond mere physical appearances-physical and physiological processes, and surface patterns-to seek deeper meanings and causes by paying attention to the meaning and motivation behind actions. But hermeneutics was not a barrier to objective knowledge. Hermeneutics was introduced in human geography in the mid-1970s through the writings of Buttimer (1974) and Tuan (1974), and was in late 1970s visible under two different titles: humanistic geography, and critical theory. The latter was proposed by Gregory (1978) with a view to linking the hermeneutical approach to a critique of traditional historical and regional geography. Thus by around 1980, the supposed dualism between space and society was finally broken, since with such a linguistic/hermeneutic, the problem of geographical description became much more than a mere matter of abstract logic: It became "a question of reading and decoding", of ***"making sense*** of different places and different people". Thus, as in the case of ethnography (Geertz, 1983), in geography also, the stress was now once again on local knowledge, which "requires us to attend not only to the theoretical categories which inform our accounts but also to the textual strategies through which they are conveyed".

The linguistic turn to the study of human geography through the introduction of hermeneutics now made geography open to experimentation with different modes of representation, so that the discipline reopened its doors to cultural studies through greater interaction across the frontier with other branches of cultural studies. Thus was completed the postmodernist journey in human geography, beginning in the 1970s, through political economy of the disorganization of capital (and thereby highlighting the relationship between history and geography, structure and agency), followed by the dissolution of the barrier between the social and the spatial, arrived at through revival of the frontier between geography and sociology by giving greater attention to social theory. This was followed by opening of the frontier with other disciplines focused on the study of cultures. The ***post-modern*** journey thus followed a route in an order that represented the reversal of the route charted by ***modern*** geography (Fig. 14.2), but the reverse journey was imposed by the turn of developments in the contemporary capitalistic system-a consequence of history. It represents essentially a process of evolution so that post-modernism must not be seen as a negation of everything that earlier human geography represented, but as "a commentary upon it" (Gregory, 1989). The crisis of "modern" geography until the mid-1970s lay in that geography had cut itself loose from mainstream philosophy, humanities and the social sciences so that the postmodernist response lay in a realignment with these other streams of knowledge for intellectual leads, particularly with social theory, but asserting at the same time that there is no scope for grand theory in human geography. Indeed, post-modernism focuses on the creative tension between the different theoretical formulations, so that post-modernism is essentially polyphonic: It combines a number of individual but harmonious melodies, i.e., points of view or perspectives, with a view to arriving at more comprehensive explanation of social reality.

Gregory (1989) has identified three distinguishing features of postmodernist thought in geography: First, space-time specificity in social explanation. This implies that post-modernist thought in social science insists on the understanding of "a world which is meaningful for the people who live within it". Reflexiveness (implying focus on the role of the human agent's action upon himself or the identity between object and subject) is intrinsic to post-modernist work in the social sciences but this does not imply complete repudiation of the scientific principles of naturalism though the search for universal theories is eschewed *(see:* Gregory, 1991). Secondly, post-modernism insists on distancing itself from the totalization (i.e., the concept of the society as a totality following a universal process of historical change irrespective of space context). Post-modernist thought holds that: "The ebb and flow of human history is not reduced to the marionette movements of a single structural principle, whatever its location, and the differences which make up human geographies are not explained by some central generating mechanism". Thirdly, post-modem human geography represents a *critique* of spatial science it is *not a continuation* of it-but at the same time, *it does not represent a break with the past.* Post-modem geography focuses on the essential *spatiality* of social life (rather than the geometry of social behaviour) (Dikshit 2018).

1.7 SELF-ASSESSMENT TEST

Discuss about the contribution of the Greeks and Romans during ancient times.

Discuss about the dualism and dichotomies in geography.

Discuss about the impact of quantitative revolution in geography

Discuss about the Critical Revolution in geography.

1.8 SUMMARIES AND KEY POINTS

Following points have been discussed in this section:

- General character and development of Geographic knowledge: Contributions of Greek, Roman and Indian scholars during the ancient period and Arab scholars during the medieval period;
- Different schools of Modern Geography;
- Concept of paradigm shifts in Geography;
- Dualism and dichotomies in Geography: Physical Geography and Human Geography, Regional Geography and Systematic Geography, Ideographic approach and Nomothetic approach;
- Positivism and Quantitative revolution in Geography; Hartshorne-Schaefer debate, System approach in Geography;
- Critical revolution in Geography; Humanistic Geography; Radical Geography; Behavioural perspectives in Geography;
- Welfare Geography; Feminism and Feminist Geography;
- Postmodernism and Postmodern Geography

1.9 STUDY TIPS

Fundamentals of Geographical Thought: Sudeepta Adhikari

Geographical Thought: A Contextual History of Ideas: R.D.Dikshit

Evolution of Geographical Thought: Majid Husain

Art and Science of Geography: Integrated Readings: R.D.Dikshit

Explanation in Geography: David Harvey

Syllabus

Group – GEO309T.2: Economic Geography, Transport Geography and Geography of Trade

(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

1. Scope and Advancement of Economic Geography; Concept and Classification of Resources;
2. Concept of Agricultural Region; Concept and Measurement of Agricultural Productivity and Efficiency; Green Revolution and White Revolution in India
3. Concept of Industrial Region and Industrial Complex; Special Economic Zone; Growth of IT Industry in India; Industrial Policy of India: Role of Liberalization; Privatization and Globalization
4. Concept of Digital Divide
5. Concepts and Measures of Distance, Accessibility and Connectivity; Transport Cost: Factors and Comparative Cost Advantages; Freight Corridor
6. Concept of Ring road, By-pass, Golden Quadrilateral, North-South and East-West Corridor; Significance of Trade in Regional and National Economy
7. Concept of Export Processing Zones, Exclusive Economic Zones, Forward Trading and E-commerce; Freight Equalization Policy on Indian Trade
8. Role of GATT and WTO in International Trade; Issues Related to FDI in India's Retail Sector and Cashless Economy

A. Introduction

Economic geography is the study of the location, distribution and spatial organization of economic activities across the world. It represents a traditional subfield of the discipline of geography.According to Dudley Stamp, Economic Geography 'involves consideration of the geographical and other factors which influence man's productivity, but only in limited depths, so far as they are connected with production and trade.'Economic Geography investigates the diversity in basic resources of the different parts of the world. It tries to evaluate the effects that differences of physical environment have upon the utilization of these resources. It studies differences in economic development in different regions or countries of the world. It studies transportation, trade routes and trade resulting from this development and as affected by the physical environment. This competitive attitude gives rise to socio-economic problems. Economic Geography, therefore, also aims at resolving such problems by better and efficient utilization of limited resources through rational, systematic, scientific and long-term planning.They still have remained concerned with the ideas of region and are applying regional methods of investigation to the demarcation of region.

B. Learning Objectives

After reading this SLM, you will be enabled to

- Know what isagricultural region;
- Measure the agricultural productivity and efficiency;
- The effects of Green revolution and White revolution in India
- Conceptualize Special Economic Zone;
- know Growth of IT industry in India;
- analyse the Industrial policy of India:
- interpret the role of Liberalization; Privatization and Globalization
- Conceptualize Digital Divide
- Know different measures of distance, accessibility and connectivity indices
- Interpret the Transport cost: factors and comparative cost advantages; Freight corridor;
- Conceptualize ring road, By-pass, Golden Quadrilateral, North-South and East-West Corridor; Export Processing Zones, Exclusive Economic Zones, Forward trading and E-commerce
- Interpret the Significance of trade in regional and national economy;
- Contextualize the Concept of Freight equalization Policy on Indian trade
- Interpret the role of GATT and WTO in international trade, and the Issues related to FDI in India's retail sector and cashless economy

C. Assessment of Prior Knowledge

To assess the students' prior knowledge, they may be asked

- What do you mean by economic geography?
- What are the different types of economic activities?
- What do you mean by resource?

D. Learning Activities

The learning process will involve Personal Contact Programmes, discussion, debate and interaction among students themselves, and students and teacher. During the Personal Contact Programmes, students may be assigned to prepare assignments on the issues of optimum population, deep ecology, criticism of the concept agricultural regions, green revolution, white revolution, special economic zone, digital divide etc.

E. Feedback of Learning Activities

Once the learning process is completed, internal assessments will be conducted. On the basis of evaluation reports of the internal assessments, some areas of the syllabus will be refocused depending on students' requirements.

Unit - 1

2.1 Scope and Advancement of Economic Geography

Economic Geography is the study of human beings and their economic activities under varying sets of conditions. Geographers are of different opinions as regarding the definition of the subject.

In fact, different authorities have defined Economic Geography in a variety of ways but their opinions converge at a common point of accord, where it means the study of the spatial distribution of human's economic activities in relation to its environment, be it physical or non-physical.

According to Dudley Stamp, Economic Geography "involves consideration of the geographical and other factors which influence man's productivity, but only in limited depths, so far as they are connected with production and trade."

Professor E. W. Zimmermann pointed out that, Economic Geography deals with the economic life of man with relation to environment.

R. S. Thoman in his book 'The Geography of Economic Activity' has remarked, "Economic Geography may be defined as an enquiry into the production, exchange and consumption of goods by people in different areas of the world. Particular emphasis is placed on the location of economic activity — upon asking just why economic functions are situated where they are in this world."

J. MacFarlane describes Economic Geography as the study of "influence exerted on the economic activity of man by his physical environment, and more specifically by the form and structure of the surface of the land, the climatic conditions which prevail upon it and the spatial relations in which its different regions stand to one another."

In the words of Hartshorn and Alexander: "Economic Geography is the study of the spatial variation on the earth's surface of activities related to producing, exchanging and consuming goods and services. Whenever possible the goal is to develop generalizations and theories to account for these spatial variations."

Surpassing all, Chisholmes says that Economic Geography is presumed to "form some reasonable estimate of the future course of commercial development," as determined by geographical factors.

Scope of Economic Geography

We may consider the Earth as the abode of human and its resources are their legacy. Being most dynamic, human being is never satisfied with mere living. They have always tried to refine theirliving conditions and environment. They are never satisfied with the simple food, nature has provided him; they have devised ways for preparing food.

Their shelters are not merely designed for simple protection, but should also be comfortable in every aspect and must match with the modern style. In fact, people satisfy not only their physical needs but also their cultural needs.

These inclinations or intentions of human mind have led to the exploitation of the earth's resources in a number of ways down from days of Palaeolithic society till the present time, but always within certain limits imposed by Nature. One unique feature of man is that, he understands the laws that govern the functioning of Nature and makes use of them in his own way of life.

The study of the manner of exploitation of the earth's resources and the limits set by physical environment is the proper scope of Economic Geography. It 'deals with the productive occupations and attempts to explain why certain regions are outstanding in the production and exportation of various articles and why others are significant in the importation and utilization of these things'.

In this study of interdependence of production, emphasis should be given upon the degree of human initiative and the nature of physical forces enacting to shape certain life-patterns. They should be studied not in isolation but as a comprehensive system of interaction between man and Nature.

However, it is not content only with the analysis of the present pattern of productive occupations, it also studies their dynamics, for global resources change not only in response to increasing knowledge, improved skills and techniques, but also, perhaps more importantly, in relation to changing socio-political objectives. Thus, Economic Geography is a much-embracing subject.

It not only aims at the understanding of different natural phenomena but also takes cognizance of racial traits and customs, advantages of an early start, availability of capital and labour, accumulated technical knowledge and skilled management, stability of governments, government aids or hindrances in the form of tariffs, subsidies or urbanization schemes and so on.

The fundamental differences in the life-styles of various societies in different parts of the world largely stem from the diversity in the physical environment, especially climate. Climatic condition differs quite distinctly from one region to another with the resultant differences in human needs. People living in cold countries, thus, require warm clothes; those in hot countries require scanty and light clothing.

People of monsoonal countries of South-East Asia take rice and fish as their staple food, those in temperate regions prefer wheat. The inhabitants of the temperate regions are more energetic and industrious than those of the warm tropical countries. Such differences in man's basic life patterns can be explained only in terms of their varying natural conditions.

Initially, at the dawn of human civilization, man's needs were certainly very limited and so easily supplied by his habitat. Even at present, a primitive man's needs remain few. He satisfies his needs by the articles which are easily obtainable from his immediate surroundings. In contrast, a 'civilized' man's needs are great and complex. They cannot be satisfied near-at-hand; they need to be supplemented from far and wide.

In fact, none of the modern countries of the world are self-sufficient. The civilized man, therefore, depends a lot on the supplies of far-away regions. This gives rise to commerce. So, we may comment that the function of Economic Geography is to study the manner in which trade and commerce are related to the earth on which they are transacted.

'Thus, Economic Geography investigates the diversity in basic resources of the different parts of the world. It tries to evaluate the effects that differences of physical environment have upon the utilization of these resources. It studies differences in economic development in different regions or countries of the world. It studies transportation, trade routes and trade resulting from this development and as affected by the physical environment.

The problem of economic resources has become more complex today with millions starving and unemployed. Such problems are more acute in the countries and among people who believe in material rather than spiritual progress. A man born and brought up under western civilization believes in creature comforts. He tries, by all means, to improve his life-style which is based on competition.

This competitive attitude gives rise to socio-economic problems. Economic Geography, therefore, also aims at resolving such problems by better and efficient utilization of limited resources through rational, systematic, scientific and long-term planning.

Humboldt, a famous 19th century German geographer, remarked that, 'the diversified riches of the earth are a vast source of human enjoyment, and, therefore, man's highest development requires that we put these riches into a common world stream of understanding and use.' This can only be achieved through the study of Economic Geography.

H. H. McCarty has aptly remarked:

Economic Geography concerns with the solution of economic problems.Hence, the student of Economic Geography must be trained along four principal lines:

a) Recognize problems and state them in a manner.
b) Develop hypotheses that promise solutions to those problems.
c) Test the adequacy of these hypotheses in providing solutions for these problems.
d) Relate tested hypotheses to other generalizations in the body of theory.

In this way Economic Geography contributes to international understanding. It, undoubtedly, broadens our knowledge and outlook to a great extent and enables us to acquire a humanistic view-point. It is essential for the liberalization of our education system for the future citizens of the modern world, so that, with its help and under its influence, he might work for true global understanding between different nations.

In pursuit of that goal the economic geographer asks three basic questions:

(i) Where is the economic activity located?
(ii) What are the characteristics of the economic activity?
(iii) To what other phenomena are the economic activity related?

According to Loyod and Dicken (1972), "As a behavioural science with spatial dimension of economic system, economic geography is concerned with the construction of general principles and theories that explain the operation of the economic system."

A major change in the study of economic geography has been in the form of behavioural approach and systems analysis. Behaviour denotes the actions of individuals or groups, it follows that economic phenomena in some way reflects individual and group values, policies and decisions.

While system is a set of identified elements so related together that they form a complex whole. Systems analysis means considerations of such a whole as a whole, rather than as something to be analysed into separate parts.

Economic geographers utilise the system concept in order to better understand the component elements of some part of reality, and the relations between them. In fact, economic geography now has grown as a developed branch of geography with specialised branches having their own status and importance.

Economic geography is sometimes approached as a branch of anthropogeography that focuses on regional system of human economic activity. Study may focus on production, exchange, distribution and consumption of item of economic activity. Allowing parameter of space time and item to vary, a geographer may also examine the flow of material, commodity, population and information from different parts of the economic activity system.

Thematically economic geography can be divided into these sub-disciplines:-

- Geography of Agriculture
- Geography of Industries
- Geography of International Trade
- Geography of Resources
- Geography of Transport & Communications
- Geography of Finance.

As the study of production, distribution and consumption of material wealth across space, economic geography covers the discursive and government creation of commodities and industries, their location and movement, and the associated effects, interactions and confrontations with people and ecosystems. Economic geographers examine issues like social relations of commodity chains (e.g. textiles from Bangladesh to the U.S.), industrial clustering (Silicon Valley), migration and remittances, corporate control of global "free trade" governance (Investor State Dispute Settlement), "global cities" as nodes of globalization, and social mobilization and solidarity across regions (e.g. Featherstone's book Solidarity).

Economic transformation proceeds differently in different places, and the mechanisms involved neither originate nor operate evenly across space. Our concern is both with the ways in which the forces making for economic change, adaptation and novelty shape and reshape the geographies of wealth creation, work and welfare, and with how the spatial structures and features so produced themselves feedback to influence the forces driving economic evolution (Boschma and Martin, 2010).

Importance of the Study of Economic Geography

The main objective of Economic Geography is, as expounded, to examine man's economic achievement in terms of production and consumption in the light of his environment. To assess the relative importance of the study of this branch of geography, we have to evaluate the purposes that it serves.

Economic Geography, fundamentally, maintains a very close relation with man's economic welfare as other social sciences do; but the approach is radically different. Through various stages of interpretation and analysis it, in the final stage, attempts to point out the potential for development of a region, occupied by a certain group of people.

Disparity in the state of economic well-being and level of production is a common phenomenon. In order to eliminate such disparity, mobilization of resources is imperative. A careful study of the situation has to be undertaken before any move is initiated to mobilize resources to resolve problems. Economic Geography accomplishes this job.

The above condition may further be elaborated. The idea of the Damodar Valley Corporation was inherited from the Tennessee Valley Authority of USA. But the DVC failed to maximize the relative benefits as the TVA did.

At the time of installation of the multipurpose river valley project in the Damodar Valley area only the aspects of physical setting were compared, ignoring totally the cultural elements like the level of technology, industrialization, capital availability etc.

Thus, due to lack of proper cultural setting, the DVC failed to reap the same degree of benefit as the TVA. Because of latitudinal locations, climatic conditions in parts of Canada are almost similar to those of the CIS. The level of technological development also bears identity.

Under this identical framework, introduction of a more cold-resisting variety of wheat, developed in either of these countries can also meet with a similar degree of success. On the other hand, the early British colonial settlers in Ceylon (presently Sri Lanka) committed a great mistake in their attempt to introduce oat, barley, and wheat cultivation in the country while both the climatic as well as the pedogenic conditions were, by no means, conducive to those crops.

Drastic transformation of the existing pattern of culture by another is not scientifically justified. A pattern of culture evolves out of the dynamic interaction between man and nature. Despite the physical identity between two countries, culture of one country may not fit that of another.

Stages of past growth or heritage play a very decisive role in shaping the present destiny of man. It is, therefore, important to note that the factor of heritage, be it natural or cultural or human, should not be overlooked since any enforced evolution or imposition of culture may bring disastrous results.

Economic Geography makes a comparatively humble and integrated approach to such problems. It describes a country or region in terms of its natural, human and cultural environments with relation to man's economic way of life. A certain set of geo-economic conditions led to the development of the Japanese method of cultivation.

Prior to the introduction of such productive system of agriculture in India a careful examination of India's agrarian infrastructure is necessary; otherwise, mere adoption of such system may not produce fruitful results.

Economic Geography serves the purpose of identifying the influence that the environment exerts on man through the preservation of the multiple geo-economic conditions of different parts of the world. Any attempt that aims at the balanced development of economy could not succeed without the complete understanding of the man-environment inter-relationship.

Without any such knowledge economic relationship is bound to end up in a fiasco. It is a fait accompli. Economic Geography, therefore, serves as an essential tool for reducing and finally eliminating world societies' disparity gaps by scientific study of their economic resources, modern needs and cultural heritages.

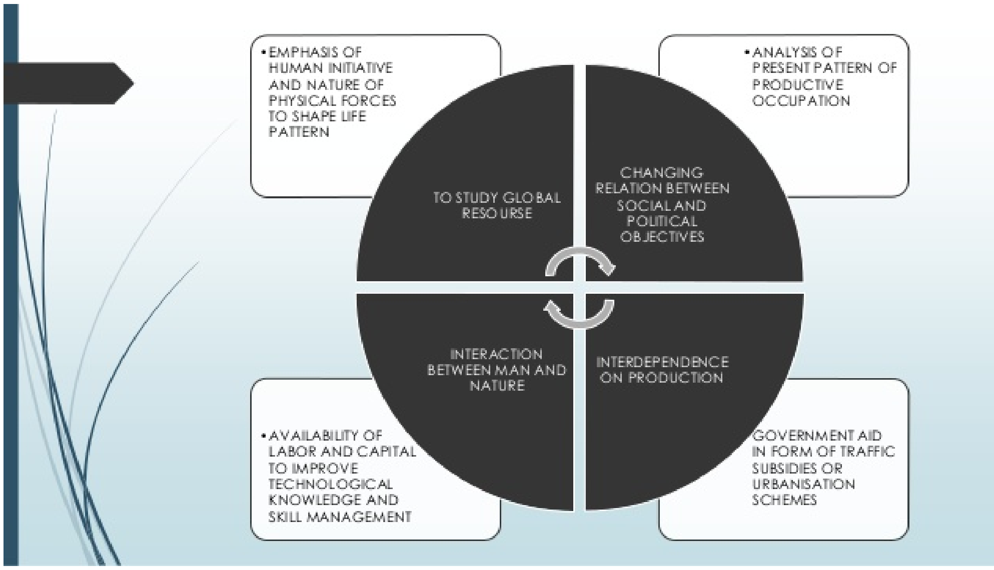


Figure: Importance of Economic Geography

Bringing Geography and Economics to the Same Table

Economists have asked why certain places grow, prosper and attain a higher standard of living at least since Adam Smith's The Wealth of Nations in 1776. Smith was motivated to understand the reasons why England had become wealthier than continental Europe. While Smith is widely considered the father of modern economics his most important theorems originated in geography. When he said that "the division of labour is limited by the extent of the market," he was referring to the geographical extension of market areas in Scotland as transport costs declined, which in turn allowed larger-scale and more geographically concentrated production, organized in the form of the factory system. The transition from artisanal production to a modern industrial economy, with a 4800 per cent productivity increase, was intrinsically geographic.

The transition that Smith analysed was profound: artisans disappeared; production become more centralized in large factories and towns, creating a geography of winning and losing places; while the incomes of industrial capitalists increased a new industrial working class faced lower incomes than artisans and more difficult working conditions. Still, there was a long-term take-off of per capita income that ended centuries of economic stagnation in the West (Maddison, 2007). Critically, Smith, and others, showed that the division of labour inside the new factories was key to the astonishing productivity gains of the factory system, but that it also picked winner and losers in terms of both individuals and social relationships and geographic places. Smith was not only concerned with the positive aggregate economic effects of the new system, but also the more complex picture of human and geographical development (Phillipson, 2010).

The processes of change that motivated Adam Smith are still at work and are no less complex or profound. Just like Smith's industrial revolution, the much-heralded 2 Knowledge Economy has created significant wealth, but the distribution of benefits is highly skewed. Indeed, there are elements of a winner-take-all tournament that favours the lucky highly skilled, with increasing income disparities. Many individuals with high levels of human capital face economic insecurity and diminished career perspectives. These dilemmas are not new: from the time that Smith wrote in the mid-18th century, through Marx's reflections of the mid-19th century, income disparities were so great that the viability the whole industrial-market (or, for Marx, "capitalist") system was called into question. The system was prone to wild swings in performance, diminished growth prospects, and deteriorating social conditions. In the 20th century these conditions spawned political instability witnessed by revolutions, and the rise of nationalism, fascism and communism. Yet in the long sweep of history, capitalism has generated the biggest boom with increases in standards of living never before imaginable for the majority of the world's population.

Even in the worst of times in the past, there were very wealthy local economies; just as in the best of times in the past, there were pockets of stagnation and poverty. The objective of this chapter is to provide a review of the intellectual history of economic geography as it relates to economic growth and economic development. We will show that economic development always has a complex interplay of winners and losers in terms of groups of people and types of places. Yet this pattern is not immutable. The less-successful people and places represent under-utilized capacities of the system. Moreover, the progress of the modern capitalist economy always begins in specific particular places; it does not spring uniformly from all territories at the same time, but diffuses from innovative places to other places across the economic landscape.

After we investigate the geographical dynamics of economic growth, this Chapter defines some new approaches to address the down-sides of the process. To do so, we will challenge some of the sacred cows of economic theory and policy to make a new meal or even a feast of future possibilities. The conventional wisdom tinkers at the margins of the growth process but does little to address the ways that the economy picks winning people and places, and under-utilizes the capacities of other people and places. By contrast, we shall show that with a deeper understanding of the geographical wellsprings of growth and development in capitalism, there are opportunities for higher growth and, most importantly, better development for both people and places.

Economists and Economic Geographers

Generally, spatially interested economists study the effects of space on the economy. Geographers, on the other hand, are interested in the economic processes' impact on spatial structures.

Moreover, economists and economic geographers differ in their methods in approaching spatial-economic problems in several ways. An economic geographer will often take a more holistic approach to the analysis of economic phenomena, which is to conceptualize a problem in terms of space, place, and scale as well as the overt economic problem that is being examined. The economist approach, according to some economic geographers, has the main drawback of homogenizing the economic world in ways economic geographers try to avoid.

Advancement of Economic Geography

Some of the first traces of the study of spatial aspects of economic activities can be found in seven Chinese maps of the State of Qin dating to the 4th century BC. Ancient writings can be attributed to the Greek geographer Strabo's Geographika compiled almost 2000 years ago. As the science of cartography developed, geographers illuminated many aspects used today in the field; maps created by different European powers described the resources likely to be found in American, African, and Asian territories. The earliest travel journals included descriptions of the native peoples, the climate, the landscape, and the productivity of various locations. These early accounts encouraged the development of transcontinental trade patterns and ushered in the era of mercantilism.

World War II contributed to the popularization of geographical knowledge generally, and post-war economic recovery and development contributed to the growth of economic geography as a discipline. During environmental determinism's time of popularity, Ellsworth Huntington and his theory of climatic determinism, while later greatly criticized, notably influenced the field. Valuable contributions also came from location theorists such as Johann Heinrich von Thünen or Alfred Weber. Other influential theories include Walter Christaller's Central place theory, the theory of core and periphery.

Fred K. Schaefer's article "Exceptionalism in geography: A Methodological Examination", published in the American journal Annals of the Association of American Geographers, and his critique of regionalism, made a large impact on the field: the article became a rallying point for the younger generation of economic geographers who were intent on reinventing the discipline as a science, and quantitative methods began to prevail in research. Well-known economic geographers of this period include William Garrison, Brian Berry, Waldo Tobler, Peter Haggett and William Bunge.

Contemporary economic geographers tend to specialize in areas such as location theory and spatial analysis (with the help of geographic information systems), market research, geography of transportation, real estate price evaluation, regional and global development, planning, Internet geography, innovation, social networks.

Approaches to Study

As economic geography is a very broad discipline, with economic geographers using many different methodologies in the study of economic phenomena in the world some distinct approaches to study have evolved over time:

Theoretical economic geography focuses on building theories about spatial arrangement and distribution of economic activities.

Regional economic geography examines the economic conditions of particular regions or countries of the world. It deals with economic regionalization as well as local economic development.

Historical economic geography examines the history and development of spatial economic structure. Using historical data, it examines how centres of population and economic activity shift, what patterns of regional specialization and localization evolve over time and what factors explain these changes.

Evolutionary economic geography adopts an evolutionary approach to economic geography. More specifically, Evolutionary Economic Geography uses concepts and ideas from evolutionary economics to understand the evolution of cities, regions, and other economic systems.

Critical economic geography is an approach taken from the point of view of contemporary critical geography and its philosophy.

Behavioural economic geography examines the cognitive processes underlying spatial reasoning, locational decision making, and behaviour of firms and individuals.

Economic geography is sometimes approached as a branch of anthropogeography that focuses on regional systems of human economic activity. An alternative description of different approaches to the study of human economic activity can be organized around spatiotemporal analysis, analysis of production/consumption of economic items, and analysis of economic flow. Spatiotemporal systems of analysis include economic activities of region, mixed social spaces, and development.

Alternatively, analysis may focus on production, exchange, distribution, and consumption of items of economic activity. Allowing parameters of space-time and item to vary, a geographer may also examine material flow, commodity flow, population flow and information flow from different parts of the economic activity system. Through analysis of flow and production, industrial areas, rural and urban residential areas, transportation site, commercial service facilities and finance and other economic centres are linked together in an economic activity system.

New Economic Geography

With the rise of the New Economy, economic inequalities are increasing spatially. The New Economy, generally characterized by globalization, increasing use of information and communications technology, the growth of knowledge goods, and feminization, has enabled economic geographers to study social and spatial divisions caused by the rising New Economy, including the emerging digital divide.

The new economic geographies consist of primarily service-based sectors of the economy that use innovative technology, such as industries where people rely on computers and the internet. Within these is a switch from manufacturing-based economies to the digital economy. In these sectors, competition makes technological changes robust. These high technology sectors rely heavily on interpersonal relationships and trust, as developing things like software is very different from other kinds of industrial manufacturing—it requires intense levels of cooperation between many different people, as well as the use of tacit knowledge. As a result of cooperation becoming a necessity, there is a clustering in the high-tech new economy of many firms.

Social and Spatial Divisions

As characterized through the work of Diane Perrons in Anglo-American literature, the New Economy consists of two distinct types. New Economic Geography 1 (NEG1) is characterized by sophisticated spatial modelling. It seeks to explain uneven development and the emergence of industrial clusters. It does so through the exploration of linkages between centripetal and centrifugal forces, especially those of economies of scale.

New Economic Geography 2 (NEG2) also seeks to explain the apparently paradoxical emergence of industrial clusters in a contemporary context; however, it emphasizes relational, social, and contextual aspects of economic behaviour, particularly the importance of tacit knowledge. The main difference between these two types is NEG2's emphasis on aspects of economic behaviour that NEG1 considers intangible.

Both New Economic Geographies acknowledge transport costs, the importance of knowledge in a new economy, possible effects of externalities, and endogenous processes that generate increases in productivity. The two also share a focus on the firm as the most important unit and on growth rather than development of regions. As a result, the actual impact of clusters on a region is given far less attention, relative to the focus on clustering of related activities in a region.

However, the focus on the firm as the main entity of significance hinders the discussion of New Economic Geography. It limits the discussion in a national and global context and confines it to a smaller scale context. It also places limits on the nature of activities carried out in the firm and their position within the global value chain. Further work done by Bjorn Asheim (2001) and GernotGrabher (2002) challenges the idea of the firm through action-research approaches and mapping organizational forms and their linkages. In short, the focus on the firm in new economic geographies is undertheorized in NEG1 and undercontextualized in NEG2, which limits the discussion of its impact on spatial economic development.

Spatial divisions within these arising New Economic geographies are apparent in the form of the digital divide, as a result of regions attracting talented workers instead of developing skills at a local level. Despite increasing inter-connectivity through developing information communication technologies, the contemporary world is still defined through its widening social and spatial divisions, most of which are increasingly gendered. Danny Quah explains these spatial divisions through the characteristics of knowledge goods in the New Economy: goods defined by their infinite expansibility, weightlessness, and non-rivalry. Social divisions are expressed through new spatial segregation that illustrates spatial sorting by income, ethnicity, abilities, needs, and lifestyle preferences. Employment segregation can be seen through the overrepresentation of women and ethnic minorities in lower-paid service sector jobs. These divisions in the new economy are much more difficult to overcome as a result of few clear pathways of progression to higher-skilled work.

Concept and Classification of Resources

Etymologically, the word 'resource' refers to two separate words — 're' and 'source' — that indicate any thing or substance that may occur unhindered many more times. The term 'resource' had no special significance till the early part of the twentieth century.

Only in 1933, when the eminent professor of economics Erich W. Zimmermann promulgated his famous **"Concept of Resource",** the idea became so popular that numerous articles and papers started pouring in the contemporary Economic Geographical literature. Urgent need was felt to identify the new concept as a separate and important branch of study.

The word resource signifies

(a) A source or possibility of assistance.
(b) An expedient.
(c) Means of support.
(d) Means to attain given end.
(e) Capacity to take advantage of opportunities.
(f) That upon which one relies for aid, support or supply.

The above characteristics vary markedly and fail miserably to produce any comprehensive universally accepted meaning of resource. However, after critical examinations and analyses all these meanings can be grouped into two, i.e., resources may help us if we are:

(a) Taking advantage of opportunity.
(b) Overcoming obstacles or resistances.

Prof. Zimmermann's inimitable definition runs: "The word resource does not refer to a thing or a substance but to a function which a thing or a substance may perform or to an operation in which it may take part, namely, the function or operation of attaining a given end such as satisfying a want. In other words, the word resource is an abstraction reflecting human appraisal and relating to a function or operation". Therefore,resource satisfies individual human wants or attains social objectives. It also refers to the positive interaction between man and nature. Man is, of course, the most important and integral part of resource creation, as he is situated in the top of the hierarchy of resource consumption. Only the satisfaction of human beings converts anything or a substance into resource.

A thing or substance is not considered as resource when it fails to give satisfaction to human beings. Proven reserves of petroleum in the midst of inaccessible terrain or in the abyss is not considered resource as they fail to yield any satisfaction to either society or individual.Geo-thermal energy in this contemporary world is considered to be the most useful resource, but, till recently, this heat-flow was not considered as resource—because man was absolutely ignorant about its uses.

Resource must possess two important properties:

(a) Function ability, and
(b) Utility.

To define anything or substance as resource, one must critically examine whether it has the property of both utility and function ability. The presence of both utility and function ability is mandatory for resource creation. For example, a bottle of poison has function ability but it has got no utility value as food. The function ability is also the function of space and time.

The resource of yesteryears may not be considered as resource today, resource considered by one country may be considered as waste product by another country, e.g., frog is considered as delicious food in Europe while it is not edible in large sections of India. Petroleum was not considered as resource until 27th August 1859, since the world's first commercial oil-well was dug at Titusville, Pennsylvania, U.S.A.

Resource and Wealth:

In day-to-day life, a common man often uses the terms resource and wealth for same purpose and meaning. Both the words signify the same expression. But, in economics and resource study, these words convey separate meanings.

Wealth, as stated by noted economist J. M. Keynes, "consists of all potentially exchangeable means of satisfying human wants". So, wealth must possess Utility, Function ability, Scarcity and Transferability. But wealth is always measurable, i.e., wealth can be expressed in terms of measuring units, like rupees.

In this manner, culture cannot be considered a wealth as it cannot be expressed by any measuring unit.On the contrary, resource may be tangible as well as intangible substances. Anything satisfying human wants can be termed resource—be it tangible or not.

Wealth is synonymous with valuables, i.e., it should be scarce while resource may be ubiquitous or abundant, e.g., sunshine, air etc.

Properties of Wealth and Resource

All wealthare resource but all resources are not wealth. Resource incorporates much more than wealth in a sense that culture, technology, innovative power, skill and different other aspects are included in the realm of resource.

Some Discarded Ideas and Popular Misconceptions about Resources:

Since time immemorial, consciousness about resource is a part of both individual and society. In fact, when human beings began community life to attain security and opulence- individuals started to gather wealth and power for future resource creation. For the three basic necessities of survival — food, shelter and clothing—man had no other option but to be aware about the resources.

Ever since Industrial Revolution (1760), the values, ethics, culture, community life, agrarian economy all received severe blow. New concepts of social welfare state, ultra-capitalism, economic colonialism affected human life. The increasing gap between 'haves' and 'have-not's enhanced internal social tensions. To combat this, emergence of socialism, communism again increased the differences between different schools of thought.

The major popular misconceptions about resource in earlier periods were:

1. Substances or tangible things like coal, copper, petroleum etc. are resources.
2. Invisible or intangible aspects — peace, culture, wisdom, policy, decisions, knowhow, knowledge, freedom—cannot be considered as resource.
3. Only natural things or substances, freely bestowed by mother earth, can be considered resource. Resource cannot be created.
4. Human populations were not considered as resource.
5. Only the quantum and magnitude of substances, not their usability or function ability and quality, were measured.
6. Resource was considered as mere 'static' and fixed asset, its mobility, dynamism and expansion ability was totally ignored.

7. The concept of resistance was totally unknown. So, matter full of resistance and no function ability was also considered as resource.

In pre-Zimmermann era, only tangible or material substances were considered resources. Different minerals like iron ore, copper, bauxite, different fuels like coal, petroleum etc. were considered resource while intangible things like peace, culture, wisdom, policy decisions etc., were not considered as resource.

In the words of Prof. Zimmermann: "………… Whereas less important invisible and intangible aspects – such as health, social harmony, wise policies, knowledge, freedom – are ignored, even though possibly these latter are more important than all the coal, iron, gold and silver in the world put together. In fact, resources evolve out of the dynamic interaction of all these factors".

In those days, widespread belief was 'Resource cannot be created'. It is a free gift of nature. Only the natural things or substances can be considered as resource. Creation, modification or extension of resource was practically unknown to that medieval world.

The role of man was grossly underestimated in earlier times. Only after the resource concept was introduced role of man in the overall resource creation process was clearly understood. In this context we can recall the legendary remarks of Prof. Zimmermann: "man's own wisdom is his premier resource—the key resource that unlocks the universe".

Considering resource a static or fixed asset was another misconception in those days. In reality, the potential ability of resource cannot be measured precisely, as, always, it may increase with improved technological advancement. Zimmermann opined that resource is as dynamic as the civilization itself.

Early geographers were totally ignorant about the property of resistance hidden within things or substances. If resourcefulness is considered as positive aspect of resource, resistance is the opposite to that of resource like assets and liabilities or profit and loss.

Resource, Resistance and Neutral Stuff

There is a total antagonistic relationship between resource and resistance—like light and shade. This inverse relationship is the key issue in the overall scheme of resource creation. Anything or any process that restricts substance becoming resource is called neutral stuff. Fertility of soil is resource but barrenness is resistance. Rain may be considered resource but flood is resistance. In the same manner, knowledge is a key resource while ignorance is the worst type of resistance.

In this connection, the concept of neutral stuff has been introduced by Prof. Zimmermann. Anything or substance, be it tangible or intangible, should be either resource or neutral stuff. If anything or substance does not contain function ability or utility value, it is termed neutral stuff.

A neutral stuff should not necessarily remain neutral forever. What is considered neutral stuff today may transform into resource tomorrow. Man's knowledge, wisdom and technological innovation may transform neutral stuff into precious resource, e.g., petroleum was not considered resource until 1859, because man was quite ignorant about its uses while, with the development of science and technology, it is now considered as a mainstay for harnessing energy.

The process of economic development is directly proportional with the rate of conversion of neutral stuff into resource. The advancement of modern civilization is synonymous with the transformation of neutral stuff into resource.

If we peep through the windows of history, it reveals that despite having enormous amount of minerals, water resources, human resources, wealth etc., some countries could not develop themselves, while others — without having any significant minerals, water etc. —because of their technology, skill, zeal, national pride and simple endeavour were able to transform their own meagre neutral stuff into resources and ultimately witnessed meteoric rise. Bihar has about double the total resources of Japan. Yet Bihar is one of the (if not the) poorest states in India, while Japan's development is unique!!So, minimization of resistance is the only way to maximize resource creation.

Functional Theory of Resources

"Resources were defined as means of attaining given ends, i.e., individual wants and social objectives. Means take their meaning from the ends which they serve. As ends change, means must change also". This statement of Zimmermann clearly states that resource creation is a function of space and time. With increasing knowledge, function of resource may enhance.

A primitive man may not be able to harness resource from a substance but a supra-animal modern man may, by his scientific Midas touch, transform such simple substance into a precious resource. To a man of animal level resistance plays a very dominant role—where nature poses obstacle for resource creation—but, to a modern Man, knowledge plays a key role to convert neutral stuff into resource.

The tropical Africa is well endowed with huge water resources. Due to backward economy and technological drawbacks, inhabitants of that region cannot convert it into energy. On the contrary, the Japanese were able to produce huge energy from far less water resources. This is because of scientific knowledge, expertise and greater economic development.

The advancement of civilization is the product of expansion of human information base. Information about minerals—coal, petroleum, iron ore, copper etc., about agriculture — HYV seeds, pesticides, insecticides etc., about manufacturing industry—the invention of steam engines, boilers, turbines, converters etc., were possible with the increasing scientific knowledge.

This increasing knowledge reduced the resistances of the natural things or substances and converted them to resources. So, Wesley C. Mitchell had aptly said: "Incomparably greatest among human resources is knowledge".

So, with the efforts of man, through the functional or operational process, resource is dynamically created. Without human effort resource cannot be created because man is the ultimate consumer of resource. Without any operational process, a thing or substance remains neutral, resource cannot be created and what is created now may be enhanced or increased with increasing knowledge. So, resource creation process is highly dynamic in nature.

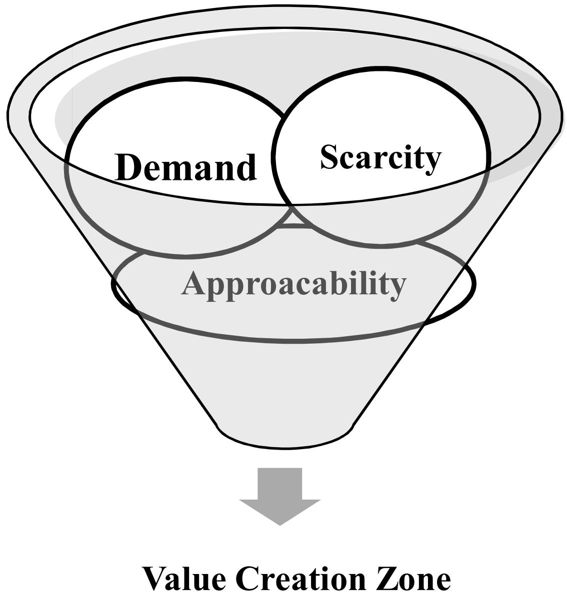


Figure 1: Creation of Resources

Dynamic Concept of Resource

Prof. Hamilton said:"It is technology which gives value to the neutral stuffs which it processes; and as the useful arts advance the gifts of nature are remade. With technology on the march, the emphasis of value shifts from the natural to the processed good".

So, resource creation process is not static, it is dynamic in nature. The thing or substance considered as neutral stuff today may be converted into precious resource tomorrow. Since the beginning of civilization, Palaeolithichuman beings started devoting his limited knowledge to convert neutral stuff into resource for his own requirement.

With the passage of time, with increasing knowledge, man was able to harness more resource from same amount of stuff. Bowman has rightly remarked: "The moment we give them human association they are as changeful as humanity itself".

With the increasing need, human beings frantically explored all possibilities or avenues to expand resource base from his existing stock. So, resource creation is a continuous and need-based operation. At the present era, when the world is passing through acute energy crisis, man is exploring possibilities to produce energy from all sources —solar energy, wind energy, geothermal energy etc.

Previously ocean current was never considered as resource but, now, people are able to convert this force into energy. So, the concept of resource is dynamic and resource study is a dynamic science.

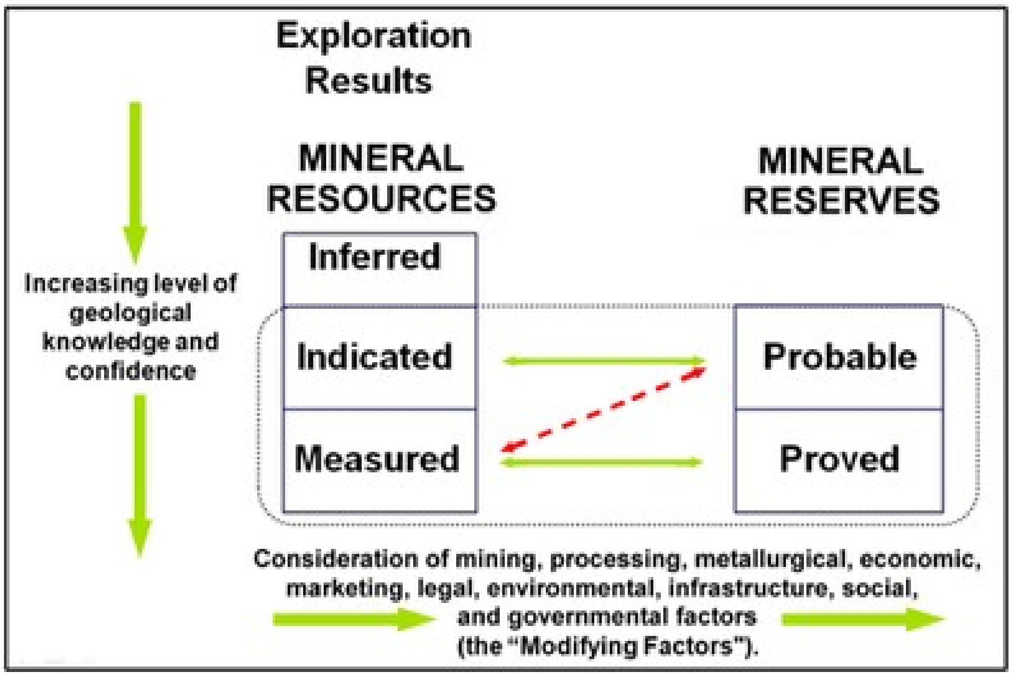


Figure 2: Dynamic Concept of Resource

Classification of Resources

The resources can be classified on the following bases:

a) On the basis of nature,
b) On the basis of durabilities,
c) On the basis of ownership,
d) On the basis of distribution

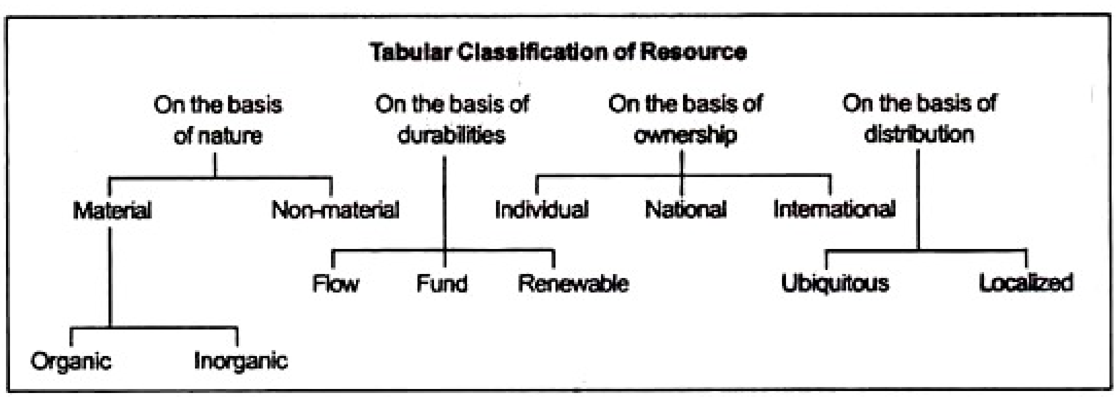


Figure 3: Classification of Resources

All those resources which have material existence, are termed as material resources, e.g. coal, petroleum, soil, air, water, plants etc. Material resources form the physical foundation of the economy; they provide essential raw materials and other commodities to support economic activity. Their use in economic activities and the related production and consumption processes have many environmental, economic and social consequences that often extend beyond the borders of individual countries or regions. The intensity and nature of these consequences depend on the kind and amounts of natural resources and materials used, the stage of the resource cycle at which they occur, the way the material resources are used and managed, and the type and location of the natural environment from where they originate. Efficient use of material resources all the way through the economy is important for assuring adequate supplies of materials to economic activities, diminishing the associated environmental burden and preventing the degradation and depletion of natural resources.

Material resources are divided into two groups:

a) Organic/ Biotic Resources,
b) Inorganic/ Abiotic Resources

All these natural objects which can regenerate and reproduce are known as biotic resources, e.g. forest, fishes, animals, human beings etc.All those things which are composed of non-living things are called abiotic resources e.g., rocks and metals.

Non-material resources, on the other hand, refer to those resources which have no material or physical existence, e.g. belief, information, wisdom, experience, technology etc. These are also known as cultural resources.

A flow resource is a resource which is neither renewable nor non-renewable, and must be used where it occurs and replenishes itself. Flow resources can only be natural resources. Humans have no influence on the process aside from collecting the resource. Examples of flow resources include solar radiation, running water, tides, geothermal, biomass, and winds.

A fund resource is a definite storage of resources uses of which gradually diminishes it. These are basically abiotic resources such as coal, petroleum, minerals etc. These generally take thousands of years to form and exist in fixed amounts in the Earth. These are also known as non-renewable resources.

Renewable resources are any resource that cycles or can be replaced within a human life span. Natural resources such as land, water, soil, plants and animals must be carefully managed, with a particular focus on how management affects the quality of life for both present and future generations. Renewable resources are seldom perfectly renewable. If their levels are heavily decreased, they may not be able to completely replenish themselves. Urban sprawl, cultivation, irrigation, grazing, deforestation, fishing, hunting, and habitat destruction can all be causes of the destruction of an otherwise renewable resource. There have been numerous efforts to prevent the mistakes that lead to the depletion of renewable resources. Despite this, destruction of renewable resources often proves to be profitable, and happens as a result. As we have become more environment-oriented in the last decade, we may hope for more reason in renewable resource management. Some authors include flow resources as renewable resources.

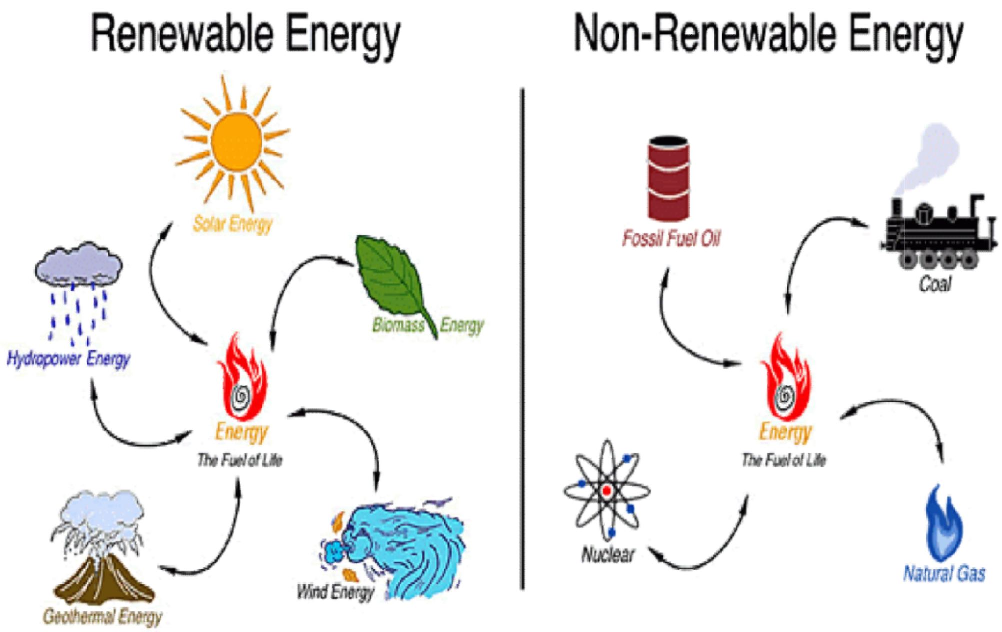


Figure: Renewable and Non-renewable Sources of Energy

Renewable	Non Renewable
1. Can be replaced by natural process in a short amount of time or can be recycled	1. These are natural resource that either cannot be replaced or may take millions of years to replace by natural process like coal and oil.
2. Can be reused or recycled and used multiple times.	2. Cannot be reused or recycled.
3. Some of the examples are: wind energy, solar power, hydroelectricity, geo thermal.	3. Some of the examples are: Petrol, coal, Natural gas, nuclear energy, fossil fuels.
4. No harm done to the environment because of its use.	4. Huge harm done to the environment because of the harmful emissions.

Advantages and Disadvantages of Renewable Energy Sources

Developing and exploiting renewable energy sources using modern conversion technologies can be highly responsive to national and international policy goals formulated because of environmental, social, and economic opportunities, objectives, and concerns:

- Diversifying energy carriers for the production fuels, electricity, and heat; enhancing energy security; and reducing the long-run price of fuels from conventional sources;

- Reducing pollution, environmental emissions, and safety risks from conventional energy sources that damage human health, natural systems, crops, and materials;

- Mitigating greenhouse gas emissions down to levels that can be sustained;

- Improving access to clean energy sources and conversion technologies, thereby helping to meet the Millennium Development Goals (MDGs) while taking advantage of the local availability of renewables;

- Reducing dependence on and minimizing spending on imported fuels;

- Reducing conflicts related to the mining and use of limited available natural resources, as most renewable energy sources are well distributed;

- Spurring economic development, creating new jobs and local employment, especially in rural areas, as most renewable energy technologies can be applied in small-, medium-, and large-scale systems in distributed and centralized application;

- Balancing the use of fossil fuels, saving them for other applications and for future generations.

Making use of renewable energy sources also has some disadvantages and drawbacks, part of which are intrinsic and part of which are due to the status of technology development:

- The spatial energy intensity (J/m^2) or density (J/m^3) of renewable energy sources is often low compared with most fossil fuel and nuclear energy sources. Consequently, space is needed where these renewable sources are converted to allow them to deliver most – and finally, perhaps all – of the energy needed. But this may create competition with other claims and requirements for the use of land, including food production, the protection of ecosystems, and biodiversity conservation. Strategies to mitigate these concerns include multifunctional land use, technologies with high conversion efficiencies, and the combination of renewables with measures to improve energy efficiency.

- Although the energy from renewable sources is most often available for free (which reduces vulnerabilities associated with the price volatility of fossil fuels), renewable energy conversion technologies are often quite capital-intensive: operating costs (fuel costs) are replaced by initial costs (installed capital costs). This can make renewables less attractive, especially when high discount rates are applied, depending on the level of investments needed as well as governmental interventions.

- The levelized cost of energy (LCOE) from renewables is often not yet competitive in the (distorted) marketplace, especially in grid- connected applications. This may change, however, as renewable energy costs are brought down through technological learning while penetrating markets. Also, using conventional energy sources will become more expensive due to resource depletion and policies to internalize external costs.

- The exploitation of renewable energy sources may entail environmental and social concerns, as experienced with, for instance, electricity production from hydropower and wind energy and the use of biomass resources.

- The intermittent character of the production of energy carriers from wind, solar, and wave energy may set specific requirements for the total energy system to achieve a reliable energy supply. It may require methods to predict renewable energy supplies many hours ahead, management of energy demands, availability of backup power, development of storage options, and enhanced flexibility of energy systems.

Why Would We Use Renewable Sources of Energy?

Today we primarily use fossil fuels to heat and power our homes and fuel our cars. It's convenient to use coal, oil, and natural gas for meeting our energy needs, but we have a limited supply of these fuels on the Earth. We're using them much more rapidly than they are being created. Eventually, they will run out. And because of safety concerns and waste disposal problems, the United States will retire much of its nuclear capacity by 2020. In the meantime, the nation's energy needs are expected to grow by 33 per cent during the next 20 years.

Renewable energy can help fill the gap. Even if we had an unlimited supply of fossil fuels, using renewable energy is better for the environment. We often call renewable energy technologies "clean" or "green" because they produce few if any pollutants. Burning fossil fuels, however, sends greenhouse gases into the atmosphere, trapping the sun's heat and contributing to global warming. Climate scientists generally agree that the Earth's average temperature has risen in the past century. If this trend continues, sea levels will rise, and scientists predict that floods, heat waves, droughts, and other extreme weather conditions could occur more often. Other pollutants are released into the air, soil, and water when fossil fuels are burned. These pollutants take a dramatic toll on the environment—and on humans. Air pollution contributes to diseases like asthma. Acid rain from sulphur dioxide and nitrogen oxides harms plants and fish. Nitrogen oxides also contribute to smog.

Renewable energy is plentiful, and the technologies are improving all the time. There are many ways to use renewable energy.Solar electricity or photovoltaic (PV) technology converts sunlight directly into electricity. Solar electricity has been a prime source of power for space vehicles since the inception of the space program.

Biomass

In recent years, the interest in using biomass as an energy source has increased and it represents approximately 14% of world final energy consumption. Estimates have indicated that 15–50% of the world's primary energy use could come from biomass by the year 2050. Many countries have included the increased use of renewable sources on their political agenda. Biomass is one such resource that could play a substantial role in a more diverse and sustainable energy mix. The energy obtained from biomass is a form of renewable energy and, in principle, utilizing this energy does not add carbon dioxide, a major greenhouse gas, to the atmosphere, in contrast to fossil fuels (Kumar et al, 2010: 2437).

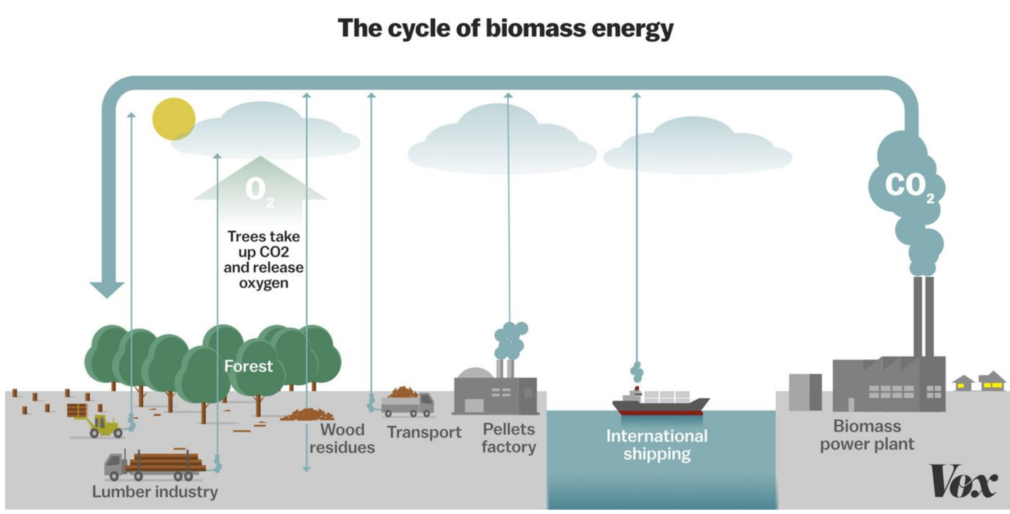


Plate: Biomass Energy as a Source of Renewable Energy

Hydropower

Hydropower is another source of renewable energy that converts the potential energy or kinetic energy of water into mechanical energy in the form of watermills, textile machines, etc., or as electrical energy (i.e., hydroelectricity generation). It refers to the energy produced from water (rainfall flowing into rivers, etc.). Hydropower is the largest renewable energy resource being used for the generation of electricity. Only about 17% of the vast hydel potential of 150,000 MW has been tapped so far. Countries like Norway, Canada, and Brazil have all been utilizing more than 30% of their hydro potential; while on the other hand, India and China have lagged far behind. India ranks fifth in terms of exploitable hydro-potential in the world. According to CEA (Central Electricity Authority), India is endowed with economically exploitable hydropower potential to the tune of148,700 MW (Kumar et al, 2010: 2437).

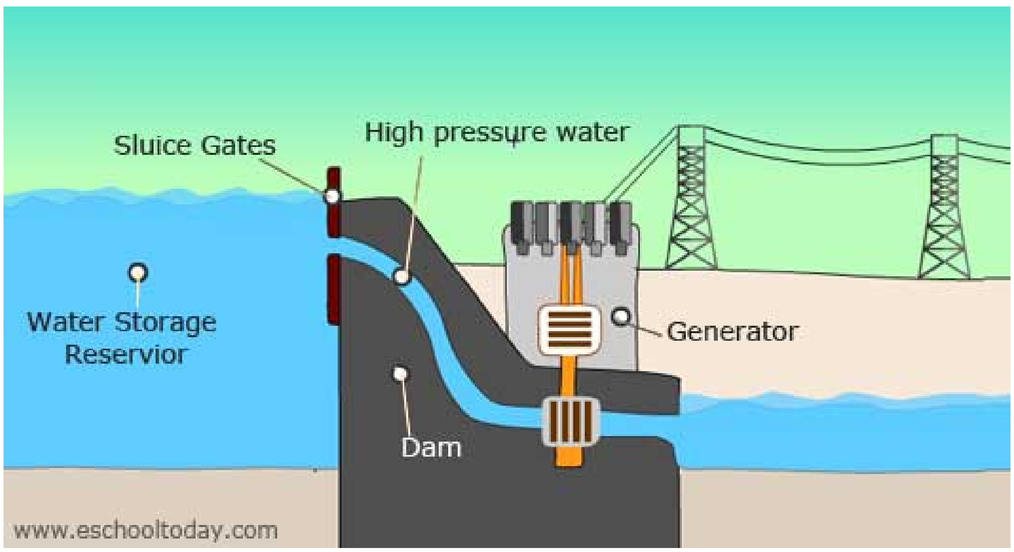


Plate: Hydropower as a Source of Renewable Energy

On the basis of ownership, resources can be classified into four classes – individual, community, national and international resources.

Individual Resources: These are owned privately by individuals. Plantation, pasture land, ponds, etc. are some of the examples of resource ownership by individuals.

Community Resources: These resources are owned by several families in a locality, e.g. a pond, a park or a garden etc.

National Resources: All the resources within the nation are called national resources. All the minerals, water resources, forests, wildlife within the political boundaries and oceanic area up to 12 nautical miles from the coast and resources within the nation, belong to the nation.

International Resources: There are international institutions which regulate some resources. The oceanic resources beyond 200 km of the Exclusive Economic Zone belong to open ocean and no individual country can utilise these without the permission of international institutions.

On the basis of distribution, resources are divided into two groups – ubiquitous and localized. **Ubiquitous** resources are those which are spread throughout the world, e.g. sunlight, wind, soil, rock etc. **Localized** resources are available only in certain places, e.g. coal, uranium, thorium, coral island etc.

Unit-2

2.2 Concept of Agricultural Region

Agriculture is the most fundamental from of human activity. An area or region with similar functional attributes is termed as agricultural system as a widen term which emphasize on the functional attributes. An agricultural system may be single farm or group of interrelated farms having similarities of agricultural attributes.

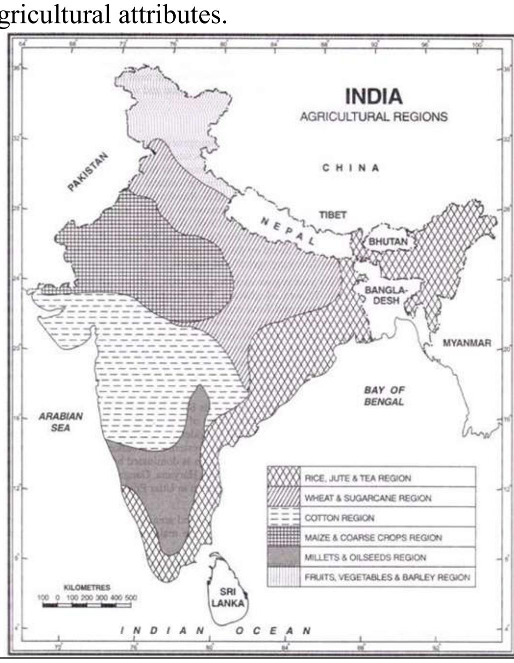


Figure: Agricultural Regions of India

Source: http://www.yourarticlelibrary.com/

The application of some statistical procedures for regional analysis has become the major concern of agricultural geographers in the recent years. Theconcept of regionalization is the process of dividingan area into territorial units of complexes of uniformitieswhich is the result of a set of processes. Regionalizationin agricultural geography is not simply anoperation of dividing the country or a region into anumber of territorial units but it is also a method ofunderstanding the agricultural pattern.An agricultural region is defined as an area having homogeneity in relief, soil type, climatic conditions, farming practices, crops produced and crop association. According to Whittlesey (1936), ‘Agricultural Region is an uninterrupted area having some kind of homogeneity with specifically defined outer limit.’

Whittlesey World Agricultural Systems/Region:

An area with similar functional attributes is as an agricultural region. The demarcation of agricultural region is also seriously constrained by the none availability of reliable data on the various aspects of agricultural patterns. The first scientific attempt for the Whittlesey in his proper. Major agricultural regions of the Earth published in 1936 in the annals of Association of American geographers (vol.26: 199-240) Whittlesey in his monumental paper delineated the agricultural system of the earth on the following five characteristics of agriculture—

(1) The crop livestock association
(2) The methods used to grow the crops produce the stock
(3) The intensity of application to the land of labor capital organization the out turn of product which results
(4) The disposal of the products for consumption
(5) The ensemble of structures used to house facilitate the faring operations

With the Boris of above indicators Whittleseyhas identified the following types of agricultural system regions—

1. Nomadic herding
2. Livestock ranching
3. Shifting cultivation
4. Rudimentary tillage
5. Intensive subsistence tillage (with paddy dominance)
6. Intensive subsistence tillage (without paddy dominance)
7. Commercial plantation
8. Mediterranean agriculture
9. Commercial grain farming
10. Commercial livestock and crop farming
11. Subsistence crop & livestock farming.
12. Commercial dairy farming
13. Specialized horticulture

A brief account of these agricultural regions is given below under separate heads.

(1) Nomadic Herding:

This is an extensive farm of animal grazing on natural pasturage involving constant on seasonal migration of the nomads their flocks nomadic handing is confined to rather sparsely populated parts of the world where the natural. vegetation is mainly grass.

Location:

Nomadic handing at present is mainly concentrated in Saharan Africa (Mauritania, Mali, Niger, Chad, Sudan, Libya, Algeria), the south western central parts of Asia the not pails of Scandinavian countries (Norway, Sweden,Finland) Northern Canada.

Characteristics:

(a) Nomadic herding ism ecological or rear ecological systems of agriculture

It is carried mainly to produce food for the family to fulfill the needs of clothing shelter recreation.

(b) It is a declining type of agriculture continues to become less important.
(c) The main characteristics of nomadic herding is the continued movement of people with their livestock in search of forage for the animals.
(d) The Bedouin of Saudi Arabia the taurag of the Sahara also practice nomadic herding in the desert semi desert areas of North Africa south west Asia.
(e) The chief characteristics of nomadic herding are described below--
    - Seasonal pattern of movement.

- Many kinds of animals grazed.
- Transhumance.

(2) Livestock Ranching:

In the extensive temperate grasslands once named by nomadic herdsmen or by hunters arc found permanent ranoles where large numbers of cattle sheep goats horses are kept.

Location:
Livestock ranching at present is manly located in the Americans, Australia, the Republic of South Africa, Brazil Argentina, Peru, New Zealand the nearest equivalent to nomadic herding is ranching.

Characteristics:

(a) The livestock ranchers specialize in animal husbanding to the exclusion of crop raising even through both live in arid on semi-arid region.

(b) Tim livestock ranchers have fixed place of residence and operate as individuals rather than with in a tribal organization.

(c) Livestock ranching differs farm nomadic herding in time important aspects.

- The vegetation cover is continuous.
- There is little or no migration.
- Ranches one scientifically managed.
- The animal’s arc rose for sale.
- Commercial grazing supports the development of town’s communications.

(3) Shifting Cultivation:

Shifting cultivation essentially this is a land rotation system. Farmers using machetes or other bladed instruments chop away the undergrowth from small pat etches of land. Then they kill the trees by cutting of a strip of bank completely around the truck. Afton the dead clean it farm the land. These cleaning techniques have gluten shifting cultivation the name or slash and burn agriculture.

Location:

Shifting cultivations the primitive form of soil utilization usually a tropical rainforests also tropical lowlands hills in the center America, Africa, and Southeast Asia, Indonesia.

Characteristics:

(a) Shifting cultivation is called by different name in different parts of the world. It is generally known as a slash burn and bush fallow agriculture. it, is variously termed as ladang in Indonesia, Milpa in central America, Mohole in the Congo central Africa.

(b) The farmer grow food only for his family in this agriculture systems. Some small surplus if any are exchanged on bartered (exchange of commodity for commodity) or sold for crash in the neigh boning markets.

(c) Shifting cultivation has been described as an economy of which the main chanaetenistues arc rotation of fields rather than rotation of crops.

(d) In the hill tracts of north-Cast India Thinning is the dominant economic activity. Our 86 percent of the people living on hill arc dependent on shifting cultivation.

(e) The shifting cultivations grow food grains-rice maize-millet- jobs-beans-vegetables- soybean.

(4) Rudimentary Tillage:

Location:

Mostly confined to the tropical lands of Central & South America, Africa, South-East Asia.

Characteristics:

(a) Crop rotation occurs most rather than field rotation.

(b) Potatoes, Sweet potatoes, Maize, Sorghum, Bananaetc are grown.

(5) Intensive Subsistence Tillage (with paddy dominance):

This from of agriculture is best developed in partially confined to the monsoon lands of Asia.

Location:

Intensive subsistence tillage dominated by paddy is practiced mostly in the tropical Asia. It is carried on mainly in China. Japan, India, Bangladesh, Myanmar, Thailand, Srilanka, Malaysia, Philippines etc.

Characteristics:

(a) Farming is also intensive that double or treble cropping is practiced. That is several crops are grown on the same land during the course of a year.

(b) Where only one crop of' paddy can be raised. The fields one normally used in the dry season to raise other food or cash crop such as sugar tobacco on oil seeds on the fiber crop jute.

(c) Asian farmer one now producing even greater yields per acre because of the recent introduction to improved varieties of hybrid rice.

(6) Intensive Subsistence Tillage (without paddy dominance):

Location: It includes interior India and North-East China.

Characteristics:

(a) Land is intensively used & worked primarily by human power.

(b) Farming in these regions suffers from frequent crop failures & famines.

(c) Wheat, Soya bean, Barley, Kaoliang crops are grown.

(7) Commercial Plantation:

The specialized commercial cultivation of cosh or estates or plantation is a very distinctive type of tropical agriculture is found many parts of world.

Location:

The term plantation agriculture was originally applied specifically to the British settlements in America then to any Inga estate in north America , west India, south-east Asia which was cultivated mainly by Negro or other colored labor.

Characteristics:

(i) A plantation is a land holding devolved to the specialized production of one tropical or subtropical crop raised for market.

(ii) Climatic hazards' strong winds, topography, drainage soil vegetation condition often handicap or many even prevent the development establishment of plantation stct. Accessibility connectivity availability of labor difficulties of clearing vegetation prevalence of discover sin sect pest , weeds, rapid deterioration of the tropical soil , soil erosion are some of the main problem of plantation agriculture.

(iii) The plantation forms arc generally large and arc found mainly in the thinly populated areas. The size of farm varies from 40 hectares in Malaya India, too 60000 hectares in Liberia. In these estates on large disciplined but unskilled labor force is necessary.

(iv) Some of the main plantation crops are rubber, oil palm, cotton, copra, beverages like coffee, tea, coco, fruits like pineapples, bananas, as well as sugar-cane jute.

(v) The continent wise analysis reveals that Asia is the leading producer of Jute (96%), rubber (90%) tea (87%) coconut (37%) tobacco (46%) of the total world production Asia share in the production of sugarcane is 39% that of banana , oil palm is 25% each (Hussain 1996).

(vi) The characteristics features of commercial plantation may be summarized as follows--

(a) Estate farming.
(b) Foreign ownership local labor.
(c) Fanning in estates is scientifically managed.

(8) <u>Mediterranean Agriculture:</u>

Within the Mediterranean climatic region where there is winter rain summer drought a distinctive type of agriculture has evolved.

<u>Location:</u>

Agricultural typology is confined to the coastal areas the Mediterranean sea in Europe, Asia, Minor, North African coastal strip. Outside the Mediterranean cost this system is found in California (USA). Central Chile the south-cast of cape province (South Africa) South-West of Western Australia.

<u>Characteristics:</u>

(a) This type of farming is also found in irrigated semi-descent descent areas in similar latitudes.

(b) The agricultural landscape of Mediterranean region has been largely affected by long day summers occurrence of rains during the winter season devices for artificial irrigation during drought periods of summers.

(c) Traditional Mediterranean agricultural is bored on what barely cultivation in the rainy wastes season raising drought resistant vine tree crops like the grape olive, fig, small livestock herding particularly of sheep goats pigs.

(d) In recent times farmer have begun using irrigation ill 1 major way which has led to the expansion of crops such as the citrus fruits.

(e) The Mediterranean land in fact the Orchard lands of the world the hart of the worlds wine industry.

(9) <u>Commercial Grain Farming:</u>

Commercial grain farming is another market oriented type of agriculture in which farmers specialize in growing wheat on less frequently rice or corn.

<u>Location:</u>

Great wheat belts stretch through Australia the plains of interior North America, the steppes of Russia, the pampas of Argentina, together the United states, Canada farmer, Soviet union.

<u>Characteristics:</u>

(i) The commercial vain fanning is basically extensive. The main characteristics of these systems are--

(a) Big farm size
(b) Comprehensive use of heavy machines
(c) Low use of irrigation fertilizer
(d) Low production rate
(e) Long distance of farm from market.

(ii) Widespread use of machinery enables commercial grain farmer to operate on this large scale indeed plan ting harvesting grain is more completely mechanized than any other form of agriculture.

(iii) Wheat is the main crop; Mize, Barleys, oat ore another important crops. The wheat production regions are divided into two belts.

(a) Winter wheat belt.
(b) Spring wheat belt.

(10) <u>Commercial Livestock & Crop Farming:</u>

<u>Location:</u>

It is found throughout Europe from h eland in the West through central Europe to Russia. It is also found in north America cast at 98'o meridian in the pampas or Argentina, Southeast Australia, Australia, South Africa, New Zealand.

Characteristics:

(a) The main characterized of the mixed farming arc that farms produce both crops livestock the tow enterprises interwove a integrated.

(b) Mixed farms one characterized by high expenditure on machinery farm building extensive use of machinery arm buildings fertilizers also by the skill experts of farmers who need to know about all aspects of farming to grow market their range of product successfully.

(c) Mixed farming is essentially associated wilt the density populated. Urbanized industrialized societies dependent.A upon high incomes for the sale or its products upon manufacturers industry for the provision of its inputs.

(d) In mixed farming a number of crops one grown. Crenels dominate the cropland use the leading grain varying with climate soil. A large portion of cereals is fed to animals on the farms on sold to Maim factures of feeding stuff. Livestock feed on crop grown on the farm graze the post wire.

(11) Subsistence Crop & Livestock Farming:

Location:

Northern Europe, Middle East, Mountain region of Mexico.

Characteristics:

(a) Produced crops & raised livestock mainly used for own subsistence.
(b) Traditional way of farming.
(c) Seeds are poor quality & animals are poorly husbanded.
(d) Capital input is normally unknown.
(e) Wheat, Maize, Rye, Barley etc are the main crops.
(f) Sheep and Goats are the most important animal.

(12) Commercial Dairy Farming:

Location:

The rearing of the cattle for milk, milk products (butter, cheep, condensed, dried milk etc) is known on daily farming. It is mainly practiced in Europe, Northern USA, Canada, Australia, New Zealand, Denmark, Netherlands, Belgium, Finland, France, and Switzerland. It contributes 40% of agricultural income.

Characteristics:

(i) Dairying is capital intensive farming. A modern dairy farm needs long here's from the farming huge amounts for the development of infrastructural facilities capital is required for the punch ore of mechanical equipments like milking machines milk freezers, feeding towers, born silos for the storage of fodder for winters.

(ii) The size of cattle in dairies varies from country to country from farm to farm depending on the size of holding. In the United Kingdom for example the ratio of cattle pasture is one cow after one acre. The average size of dairy cattle in north-west Europe is only five cows per farm.

(iii) Nearly 80% of the total milk production of the world is produced in Europe, Russia, Anglo America, Australia, New Zealand (Hussain 1996).

(iv) Modern method of daily farming cattle breeding herd management allow high yields of dairy products. A cow in temperate latitudes under normal healthy conditions yield or much as 3000kg of milk pen year.

Example: The countywide position in some the denying countries of milk pen cow pen annum have given following table.

Country	Amount (kg)
Canada	2896
Denmark	3946
Japan	4284
Netherlands	4202
UK	3797
USA	3767

Source: Production year book 91-94

It will be seen from data that the average production of milk per cow pen annum varies between 2896kg in Canada to 4284kg in Japan.

(13) Specialized Horticulture:

Specialized cultivation of vegetables, fruits, flowers is called horticulture.

Location:

Horticulture is well developed in the densely populated industrial districts of north-west Europe, Britain, Denmark, Germany, Netherlands, France, Italy.

Characteristics:

(i) In horticultural the farms arc small such farms arc located where communication links the consumption centers arc appreciably good. The land fruits' vegetable gardening is very intensively cultivated. Soil fertilizers Mimi oldie work is done by hand labor.

(ii) The market gardens are scientifically managed to achieve optimum yields hand some returns.

(iii) The important vegetation region are California Rio Ground boring of Texas Florida Netherlands, Rhone valley etc. fruits regions are west of Paris , Rhine valley lake region of Switzerland, Mendoza, Sanjuam of Argentina etc. (grapes production) south western Germany (apple).

(iv) Besides south Arab, Iraq for date. India, South-East Asia for spices, pineapple, mango etc.

Conclusion:

In this modern period, most of people depend on agriculture for food. But at this stage, many portion of the world suffers from food deficiency. So, a modern cultivation method should apply for achieving sufficiency in food.

2.3 Concept and Measurement of Agricultural Productivity and Efficiency

The productivity of a region's farms is important for many reasons. Aside from providing more food, increasing the productivity of farms affects the region's prospects for growth and competitiveness on the agricultural market, income distribution and savings, and labour migration. An increase in a region's agricultural productivity implies a more efficient distribution of scarce resources. As farmers adopt new techniques and differences, the more productive farmers benefit from an increase in their welfare while farmers who are not productive enough will exit the market to seek success elsewhere.

As a region's farms become more productive, its comparative advantage in agricultural products increases, which means that it can produce these products at a lower opportunity cost than can other regions. Therefore, the region becomes more competitive on the world market, which means that it can attract more consumers since they are able to buy more of the products offered for the same amount of money.

Increases in agricultural productivity lead also to agricultural growth and can help to alleviate poverty in poor and developing countries, where agriculture often employs the greatest portion of the population. As farms become more productive, the wages earned by those who work in agriculture increase. At the same time, food prices decrease and food supplies become more stable. Labourers therefore have more money to spend on food as well as other products. This also leads to agricultural growth. People see that there is a greater opportunity to earn their living by farming and are attracted to agriculture either as owners of farms themselves or as labourers.

However, it is not only the people employed in agriculture who benefit from increases in agricultural productivity. Those employed in other sectors also enjoy lower food prices and a more stable food supply. Their wages may also increase.

Agricultural productivity is becoming increasingly important as the world population continues to grow. India, one of the world's most populous countries, has taken steps in the past decades to increase its land productivity. Forty years ago, North India produced only wheat, but with the advent of the earlier maturing high-yielding wheats and rices, the wheat could be harvested in time to plant rice. This wheat/rice combination is now widely used throughout the Punjab, Haryana, and parts of Uttar Pradesh. The wheat yield of three tons and rice yield of two tons combine for five tons of grain per hectare, helping to feed India's 1.1 billion people.

The measurement of production and inputs required for the production of that output is known as agricultural productivity. In other words, it is an input-output ratio.

In traditional measurement of agricultural productivity geographers and economists used to take into account the inputs like labour and capital and see them as costs which are incurred in the production of agricultural produce.

The traditional approach of measurement of agricultural productivity, however, does not take into account of social and environmental costs which are also incurred in the production of crops and raising livestock.

At present, in the measurement of agricultural productivity, the question of sustainability of soil, health of ecosystem and social acceptability have become increasingly important. Agricultural productivity of a micro or macro region is closely influenced by a number of physical (physiography, climate, soil, water), socioeconomic, political, institutional and organizational factors.

Thus, agricultural productivity is a function of interplay of physical and cultural variables and it manifests itself through per hectare productivity and the total production. Agricultural productivity also depends on attitudes of the farmers towards work and their aspirations for better standard of living.

The measurement of agricultural productivity helps in knowing the areas that are performing rather less efficiently in comparison to the neighbouring areas. By delimiting the areas of low, medium and high productivity, agricultural plans may be formulated to remove and minimize the regional inequalities. It also provides an opportunity to ascertain the ground reality, the real cause of agricultural backwardness of a tract/area or region.

In the recent decades geographers and economists have developed sophisticated tools and techniques to determine the agricultural productivity.

Some of the well-known techniques developed and used for the measurement of agricultural productivity and agricultural efficiency per unit area/per unit of time are given below:

1. Output per unit area.

2. Production per unit of farm labour.

3. To assess agricultural production as grain equivalents (Buck, 1967).

4. Input-output ratio (Khusro, 1964).

5. Ranking coefficient method (Kendall, 1939; Stamp, 1960; Shafi, 1990).

6. Carrying capacity of land in terms of population (Stamp, 1958).

7. Giving weight to the ranking order of the output per unit area with the percentage share under each crop (Sapre and Desh- pande, 1964; Bhatia, 1967).

8. Determining an index of productivity (Enyedi, 1964; Shafi, 1972).

9. Computing the crop yield and concentration indices ranking coefficient (Jasbir Singh, 1976).

10. involving the area, production and price of each cultivated crop in each of the constituent areal units of the region, and then relating the out-turn in terms of money of the unit to the corresponding productivity of the region (Husain, 1976).

11. To assess agricultural production in terms of money.

12. Assessing the net income in rupees per hectare of cropped area (Jasbir Singh, 1985).

Each of the techniques advocated and applied for the measurement of agricultural productivity suffers from one weakness or the other. The application of a technique may give satisfactory results at the micro or meso level but the same technique fails to deliver the goods at the national or global level.

The input and output ratio technique seems to be a reasonably good one but the determination of inputs including environmental and social costs involved in the production is not an easy task.

The conversion of production of all crops in terms of money is also a useful technique but it is constrained by the prevailing prices of agricultural commodities which fluctuate from one areal unit to another and from one region to another.

Productivity Index = {(Production value in money of all the crops in an areal unit/ Total cropped area in an areal unit) ÷ (Production value in money of all the crops in the region/ Total cropped area in the region)} × 100

Efficiency Index = $lyn = (y1 / y) \times 100$

Where, lyn= percentage yield of crop n

y1= yield of individual crop in an areal unit.

y = yield of individual crop in the total area.

$Ei = (ly1c1 + ly2c2 + ly3c3 + ... + lyncn) / (c1+c2+c3+.....+cn)$ Where, Ei = Agricultural Efficiency Index (AEI) ly1, ly2, ly3..... lyn = The indices of different crops. c1, c2, c3.....cn = Percentages of crop area to total cropped area.

5.0. Green revolution and White revolution in India

2.4 Green Revolution

The Green Revolution in India was an effort to increase agricultural production in India via a package of industrial agriculture technologies, such as hybrid seeds, fertilizers, pesticides, and irrigation.The green revolution was a planned intervention expected to raise cereal-grain production in Third World countries. Use of high yield varieties (HYVs) was to reduce dependence on food imports and bring about food self-sufficiency. Early results were increased overall cereal-grain production, but unforeseen and undesirable social and economic consequences.

Origins of the Green Revolution

In April of 1969, then president of the Rockefeller Foundation, George Harrar called ameeting to address the problem of world hunger. In attendance were 16 leaders from the world'smajor foreign assistance agencies that were also concerned with agricultural development. Oneof those in attendance was Lowell S. Hardin, author of "Meetings that Changed the World:Bellagio 1969: The Green Revolution" (Hardin 2008). The Rockefeller Foundation had already been working with partners in developingcountries to develop technology to increase food production. They held the meeting in Bellagio,Italy to emphasize the importance of scientific advancements in farming techniques as opposedto food shipments to poor countries by aid organizations. They presented the positive impactsachieved by the major international agricultural research centres. One of these achievements wasthe introduction of new varieties of seeds developed by plant scientist Norman Borlaug that were"stocky, disease-resistant, fast-growing and highly-responsive to fertilizer" (Hardin 2008: 471).

The 1969 Bellagio conference was the impetus for the world's agricultural-development organizations' mobilization of plans for an increase in food production, later termed the "GreenRevolution". The Green Revolution quickly spread through the developing world, including thestates of India. (Hardin 2008).There were worries at the meeting about the potential for unintended consequences.These included the possibility of widening the gap between small farmers and large landowners,as well as issues of depleting soil and water resources (Hardin 2008). This paper investigates these concerns and the effects on small farmers in India by thetechnologies put forth by the Green Revolution. To accomplish these goals this literature reviewwas conducted by analysis of scholarly journal articles as well as books, websites and magazinearticles pertaining to the issue (Sebby, 2010).

Objectives

The main objective of Green Revolution introduced in 1960s was to increase the production of food grains in India by using improved technology and make India self-sufficient in food grains.

Major Inputs

- The most important strategy followed in green revolution is the application of high yielding variety (HYV) seeds. Most of these HYV seeds are or dwarf variety (shorter stature) and matures in a shorter period of time and can be useful where sufficient and assured water supply is available. These seeds also require four to ten time more of fertilisers than that of traditional variety.
- Use of pesticides
- Consolidation of holdings
- Land reforms
- Improved rural infrastructure

- Supply of agricultural credit
- Use of chemical or synthetic fertilizers
- Use of sprinklers or drip irrigational systems
- Use of advanced machinery
- Use of vector quantity

Strategies

Accordingly, in 1960, from seven states seven districts were selected and the Government introduced a pilot project known as Intensive Area Development Programme (IADP) into those seven districts. Later on, this programme was extended to remaining states and one district from each state was selected for intensive development. Accordingly, in 1965, 144 districts (out of 325) were selected for intensive cultivation and the programme was renamed as Intensive Agricultural Areas Programme (IAAP). During the period of mid-1960s, Prof. Norman Borlaug of Mexico developed new high yielding varieties of wheat and accordingly various countries started to apply this new variety with much promise. Similarly, in the kharif season in 1966, India adopted High Yielding Varieties Programme (HYVP) for the first time.

This programme was adopted as a package programme as the very success of this programme depends upon adequate irrigation facilities, application of fertilizers, high yielding varieties of seeds, pesticides, insecticides etc. In this way a new technology was gradually adopted in Indian agriculture. This new strategy is also popularly known as modern agricultural technology or green revolution.

Success

As a result of the Green Revolution and the introduction of chemical fertilizers, synthetic herbicides and pesticides, high-yield crops, and the method of multiple cropping, the agricultural industry was able to produce much larger quantities of food. This increase in productivity made it possible to feed the growing human population.

One person who is famous for his involvement in the Green Revolution is the scientist Norman Borlaug. In the 1940s, Norman Borlaug developed a strain of wheat that could resist diseases, was short, which reduced damage by wind, and could produce large seed heads and high yields. He introduced this variety of wheat in Mexico and within twenty years the production of wheat had tripled. This allowed for the production of more food for people in Mexico and also made it possible for Mexico to export their wheat and sell it in other countries. Norman Borlaug helped introduce this high-yield variety of wheat to other countries in need of increased food production, and he eventually won a Nobel Peace Prize for his work with developing high-yield crops and for helping prevent starvation in many developing countries.

In addition to producing larger quantities of food, the Green Revolution was also beneficial because it made it possible to grow more crops on roughly the same amount of land with a similar amount of effort. This reduced production costs and also resulted in cheaper prices for food in the market.

Growth in Yield of Crops in Post-Green Revolution Period

(Percent Per annum)

Crops	1967-68 to 1980-1981	1980-81 to 1989-90	1990-1991 to 1999-2000
Rice	1.46	3.19	1.27
Wheat	2.61	3.10	2.11
Coarse cereals	1.57	1.62	0.08
Total cereals	1.78	2.90	1.58
Pulses	0.66	1.61	0.96
Total food grains	1.54	2.74	1.52
Oilseeds	0.77	2.43	1.25
All Principal crops	1.43	2.56	1.31

Source : Agricultural Statistics at a Glance (various Issues), Directorate of Economics and Statistics, Ministry of Agriculture, Govt. of India.

Role of the Green Revolutions in Economic Development

At the beginning of economic development the agricultural sector is large. A large share ofpopulation depend their livelihood on agriculture and related activities. They are poor and the shareof their household expenditures for food and beverages (Engel's coefficient) is usually very high;around 70 percent. Under such a situation, even if the government tries to promote industrialization(especially heavy industrialization) with neglecting the agricultural sector, it tends to fail because ofthe lack of the market for non-agricultural sectors. Note that export-oriented industrialization is moredifficult and entrepreneurs should at first depend on the domestic market which is more familiar tothem before going to exploit export market. In this sense, the existence of the domestic market fortheir products is essential when promoting industrialization. Because the majority of people live inrural areas at this stage of economic development, the key is how to raise income and alleviatepoverty in widespread rural areas. Thus the development of agricultural sector, especially staple foodsector, should come first because majority of rural population depend their livelihood on it. If raisingincome of rural population is the key, the agricultural growth should be led by productivity growth, rather than by horizontal expansion of farmland.

Actually, as we had seen before in this paper, India had to pay a huge cost for the negligenceof agricultural sector before the mid-1960s, in the form of the lost decade from the mid-1960s tothe mid-1970s.In conclusion, the second Green Revolution in India during the 1980s was able to play acritical role in preparing a wide market in rural areas for non-agricultural products and services,which became the basis of the rapid economic growth based on non-agricultural sector developmentin the country after the 1990s. The author emphasizes here the „final demand effects" of theagricultural development in the 1980s, although the author does not deny the existence and theimportance of backward and forward linkage effects of it too.

<u>Failures of the Green Revolution</u>

- Unfortunately for many farmers the cost of machinery was too much and they simply couldn't afford it, as well as the high initial outlay, money was also required for fuel and repair.
- Many very poor farmers were tenant farmers, with little money to buy even the new seeds or fertiliser that was required.
- New irrigation schemes were required to provide the reliable source of water required by the HYVs (High Yielding Varieties of rice). As well as being expensive, in some cases where inappropriate schemes were used salinization became a problem. Dam construction in some areas also resulted in the flooding of some good farming land.
- The large amounts of fertilisers and pesticides required by the HYVs also led to serious environmental problems as they entered water supplies.
- In areas where there was an increase in mechanisation, there was an increase in unemployment with fewer people needed to do the jobs that were now done using tractors etc. The consequent increase in unemployment in rural areas led to an increase in rural-urban migration with more people moving to the cities, causing urban problems. Many farmers who had tried to take on the new technologies became heavily in debt, leading to increase stress and in some instances suicide.
- The Green Revolution created wide regional and interstate disparities. The plan was implemented only in areas with assured supplies of water and the means to control it, large inputs of fertilizers, and adequate farm credit. These inputs were easily available in at least parts of the states of Punjab, Haryana, and western Uttar Pradesh; thus, yields increased most in these states. In other states, such as Andhra Pradesh and Tamil Nadu, in areas where these inputs were not assured, the results were limited or negligible, leading to considerable variation in crop yields within these states. The Green Revolution also increased income disparities: higher income growth and reduced incidence of poverty were found in the states where yields increased the most and lower income growth and little change in the incidence of poverty in other states.
- Green revolution caused wide disparities in the production of different crops. In the name of green revolution, actually wheat revolution took place. The production of rice was limited.

Plate: Use of Pesticides

This led to raise in issues regarding Green Revolution

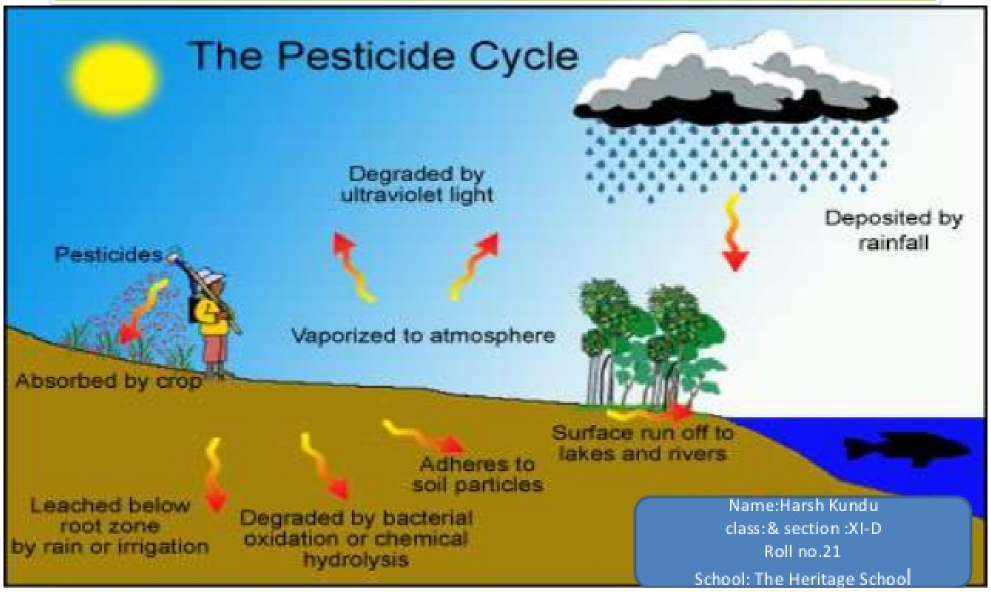


Figure: Pesticide Cycle in the Post-Green Revolution Period

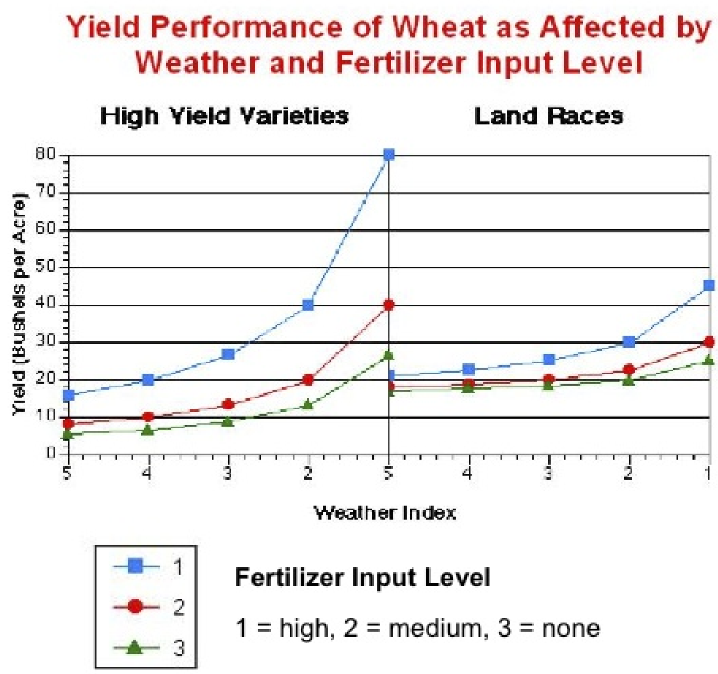


Figure: Effect of Chemical Fertilizer on Production of Wheat

Uneven Geographical Distribution of Green Revolution Technology

Thirteen North American agronomists were sent to India to assess the possibilities of technological advancement in agricultural villages. They soon realized that it would be impossible to assist all of India's 550,000 villages. Because of this the Intensive Agricultural Development Program was formed to especially select areas best suited for agricultural development. These areas would then receive the material and financial resources of the entire country for technological advancement (Shiva 1993). The country of India is comprised of a multitude of different topographies, climates and soil types, or agro-ecological zones. Because of this, the modern varieties of seeds developed to increase production were not ecologically viable for all regions. The seed types developed simply did not vary enough to be used consistently across the country (Das 1999). This provided the Intensive Agricultural Development Program with the criteria needed to select the different villages that would receive the technological packages developed by Green Revolution scientists (Shiva 1993).

Modern varieties of crops such as sorghum, millet and barley, which grew in mainly semiarid and dry land conditions were not widely developed until the 1980's, and even then there was limited availability of these seeds. The most successful and highly distributed crops were rice, maize and wheat, which needed areas that had high rainfall or sufficient irrigation (Evenson 2003). Of all of these, wheat was the most successful, causing states in the Northwest to fare particularly well because of their irrigation systems (Chakravarti 1973). Not only did this limit opportunities for farmers in arid regions, it resulted in only a portion of land suitable for growth to be utilized. This was due to the lack of irrigation facilities needed to sustain the crops, a consequence of the lack of state involvement (Das 1999), which will be discussed later on. Farmers who are able to afford irrigation systems are having problems as well. They must use expensive pumps to tap into groundwater reserves, which are depleting rapidly.

In some areas, groundwater levels are sinking up to three feet per year (Zwerdling 2009). Today, farms that had begun digging down ten feet now find themselves drilling down more than 200 feet, requiring even more expensive equipment (Zwerdling 2009). Rather than cooperating with the environment by employing traditional methods of crop rotation and the planting of diverse varieties of crops to solve the problem of uneven geographical distribution, researchers developed strains of crops that would adapt to unsuitable conditions so long as they were heavily irrigated and received applications of expensive inputs of fertilizers and pesticides (Overdorf 2009), which led to disparities between farmers who could afford these technologies and those that could not.

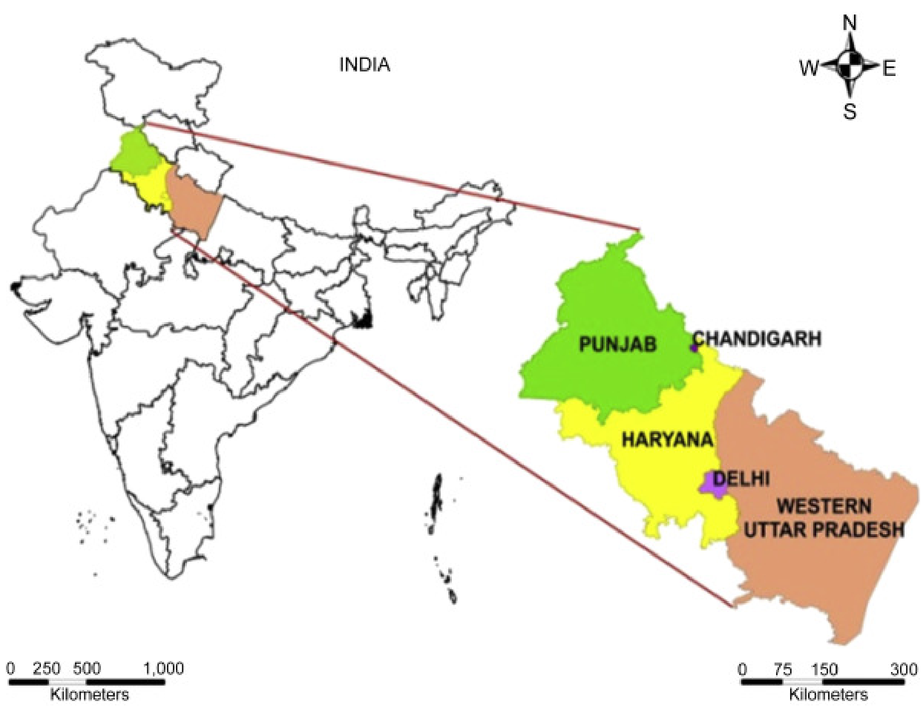


Figure: Effect of Green Revolution Limited in Few States of India

Loss of Small Farms to Large Commercial Farms

The Green Revolution came about as a means to present a solution to resource and food scarcity. However, physicist and ecological activist Vandana Shiva believes that this solution led to the "ecological breakdown in nature and the political breakdown of society [as] consequences of a policy based on tearing apart both nature and society" (quoted in Shiva 1993: 24). The Green Revolution's method to increase food production and eliminate hunger is to introduce technologies such as bio-engineered seeds and chemicals that are developed to increase yields. However, smaller farmers have fallen behind and have had to take out loans and sell land to afford the technology of the Green Revolution. These farmers were left impoverished, and there were many reported suicides by farmers too proud to beg.

According to agricultural researcher Peter Rosset, there are three important lessons we can take from the Green Revolution: First, where farmland is bought and sold like any other commodity and society allows the unlimited accumulation of farmland by a few, super-farms replace family farms and all of society suffers. Second, where the main producers of food--small farmers and farm workers--lack bargaining power relative to suppliers of farm inputs and food marketers, producers get a shrinking share of the rewards from farming. Third, where dominant technology destroys the very basis for future production, by degrading the soil and generating pest and weed problems, it becomes increasingly difficult and costly to sustain yields (Rosset 2000: 6). MonkombuSambasivanSwaminathan, who is considered the father of the Green Revolution in India, agrees that it may not have been the best plan for Indian agriculture. He attributes this to the fact that the industrialization and monoculture introduced to India by the thirteen Green Revolution have resulted in low water tables and soil that has been depleted of nutrients essential for growth.

Additionally, these techniques sparked a vicious cycle in which farmers were forced to spend more and more money on chemicals to counteract what monoculture and heavy fertilizer applications have done to their land. (Laidlaw 2008). Presently, water resources are scarce and expensive in India. Large farms benefit because they can afford canal irrigation, whereas small farmers need to resort to taking out loans with high interest rates to irrigate their fields. Keith Griffin wrote in 1979, "The growth in inequality in rural areas stem in large part from the fact that small, poor peasants who have restricted access to credit, technical knowledge and the material means of production are unable to innovate as easily or as quickly as those who are landed, liquid and literate" (Griffin 1979: 216). A survey in 1967 showed that in Punjab, the showcase state for the Green Revolution, 65% of Punjabi farmers owned fifteen acres of land or less, and their land only accounted for about 34% of the total land of the state. The rest of the land was owned by the minority of farmers who owned twenty acres or more.

While a majority of farmers in Punjab farmed ten acres or less, it was shown that only farmers that owned at least twenty acres could afford to purchase the new inputs of the Green Revolution (Newman 1997). In short, small farmers were just not equipped to keep up with the pace of the Green Revolution. To obtain credit, farmers need to have a secure tenancy or own land, which makes them more viable for investment and provides an asset on which credit can be acquired. Liquid assets are essential for purchasing commercial inputs of seed and chemicals as well as enabling a farmer the stability to hold out and wait for the most opportune time for sales and purchases. 14 Farmers who are illiterate have less access to knowledge and information regarding commercial inputs, as well as the ability to learn about proper usage and techniques. Small farmers also suffer the effects of heavy fertilization taking its toll on the land and destroying the soil. New seeds are dependent on large quantities of fertilizers, which decrease the presence of nutrients in the soil such as nitrogen, phosphorous, iron and manganese. To counteract this, farmers must apply even more fertilizers to make up for the lack of important nutrients and aid the growth of the plants (Zwerdling 2009). Furthermore, pesticides and herbicides lead to resistant species, creating further need for chemical applications. Farmers are finding it harder to stay ahead of these growing costs, but they have become dependent on the new technological inputs.

The Green Revolution favoured expensive technological investments for the "best endowed farmers in the best endowed areas, and directed away from resource prudent options of the small farmer in resource scarce regions" (Shiva 1993: 45). Because of scarce resources, competition increased among farmers--further widening gaps between social classes as well as geographical regions as discussed above. In a system where purchasing expensive equipment is necessary for a farmer to stay competitive, small farmers often must rely on unofficial creditors and money lenders. These business men loan money and charge exorbitant interest rates, sometimes as much as double the interest rate that banks charge. In the village of ChotiaKhurd, Daniel Zwerdling reported that farmers are paying interest at rates as high as 24% (Zwerdling 2009). Traditionally, agricultural knowledge and customs in India were shared among the people that had been passed down through social and cultural structures and practices. With the introduction of the Green Revolution and the commercialization of agriculture, farmers were 15 faced with a difficult shift from traditional agricultural knowledge to a more modern methodology.

Farmers now rely more on market-led fads as they are impacted by the growing competition among agri-businesses. Because of this, agriculturalists are constantly in competition--causing the need to purchase new commercial varieties of seeds, pesticides, and fertilizers. These new agricultural methods do not necessarily come equipped with the training and knowledge necessary to effectively implement them. Professor A. R. Vasavi drew on data from the India's National Sample Survey number 49 which pointed out the lack of public agencies in rural India as well as the lack of training and education to a significant portion of people (Vasavi 2005). "Only 8.4% of agriculturists had accessed information from the KrishiVigyanKendras (Farmers Science Centres); 17% received information from other agriculturists and 13% from agricultural input dealers" (Vasavi 2009: 99). Furthermore, the reports indicated that many farmers have encountered problems with their farms because of their lack of knowledge of pesticide use and commercial inputs. Soil tests must be conducted to determine the correct amount and composition of fertilizer and pesticide mixtures. A large amount of small farms have missed out completely in this process, leaving them ignorant of what they should be using on their fields (Chakravarti 1973).

In fact, the leading cause for increasing debts of agriculturists was the inappropriately large amount of pesticides they sprayed on their fields (Vasavi 2009). When this was not the case, farmers would apply too little to their land to save money, causing the fertilizers to be virtually ineffective. In other cases farmers would wait until their field was infested before buying pesticides, which also defeats the purpose of these chemicals (Chakravarti 1973). 16 In the political economy of the Green Revolution, it is evident that large, wealthy farmers in India have more influence in politics than do small farmers. Griffin made this point years ago when analyzing the Green Revolution: "The ends as well as the means of government policy usually are largely determined by the government's sources of support. In most states this support comes from local elites and economic policy is designed to further their interests . . . Seldom has a government attempted deliberately to mobilize the groups that comprise the 'wretched of the earth', i.e. Those who lack the privileges of power and the resources necessary for material progress" (Griffin 1979:176-77).

The wealthy landholders are often the creditors and political bosses that make important decisions, or they have heavy influence on those that are by providing them financial support. This makes it impossible for small farmers to share in any of the control because they lack the amount of land ownership and caste status to change their situation--when it comes down to it, the government is more prone to protect the interests of those whom they rely on to provide support when it is needed. In the 1970's and 80's a large number of small farms, unable to keep up with the competition, shut down. There were 1,375,382 landholdings in Punjab in 1971, and this number fell to 1,027,127 by 1981. This decline of nearly 25% of the farms of Punjab led to a shift toward large farms (Shiva 1993). In 1984 24% of small farmers and 31% of marginal farmers were living below the poverty line (Singh 1984). There was also a study done by the Johl Committee, who gathered in 1985 under the leadership of economist S.S. Johl to assess agricultural systems in India, that showed that while from 1977-1979 there was an increase in per hectare income, there was a decline in financial returns from farming in Punjab.

The 17 introduction of external inputs was only possible by the introduction of credit which led to the indebtedness of many farmers because of rising costs of chemicals and the necessity to apply more and more to fields as the soil continues to degrade (Shiva 1993). In 1984 there were a lot of cases in which small landowners who still technically owned their land had leased it out to wealthier farmers and then subsequently worked the very same land as a laborer, enabling them to get at least some money out of the land, since they could not afford to purchase the inputs needed for Green Revolution agriculture (Singh 1984). The increase in profitability of agriculture caused many landlords to take a new interest in farming, evicting tenants so that they may farm the land themselves. This made less land available to small farmers, as large farmers and landlords were increasingly repossessing the land for their own use. Even if they did not evict the tenant, many times landlords would charge higher rent or crop share or rotate the tenants among the different plots of land so they were not able to acquire any rights to their land (Singh 1984).

Second Green Revolution

The first Green Revolution was launched to ensure food security as there was severe scarcity of food in the country. Today, our food supply is well secure. Meeting the growing needs is within reach. Therefore, the second Green Revolution should aim at promoting sustainable livelihood, enabling the poor to come out of poverty by generating gainful self-employment. While the first Green Revolution aimed at undertaking mass agricultural production, the second Green Revolution should be to promote agricultural production by the masses. This is the Gandhian approach to equitable sharing of prosperity by involving the poor in development. For achieving this goal, we need to search for technologies which can be adopted by the farmers in arid and semi-arid regions, and those who are dependent on rainfall for crop production.

2.5 White Revolution

India's White Revolution, which has quietly swept the country during the past few decades, deserves attention equal to that given to the better-known Green Revolution. The White Revolution holds the promise of raising the nutritional status of underprivileged sections of our people. With a production forecast of 74 million tonnes in 1998-99, India has become the largest milk producer in the world. From being a major importer of dairy products in the 1950s, India has now become an exporter. Milk has become India's most important farm commodity, the value of its output (in 1994-95) of Rs 500,000 million exceeding that of paddy.

Dr.VergheseKurien – The Legend of White Revolution in India

Dr.VergheseKurien was born on November 26, 1921 at Kozhikode, Kerala. His father was a Civil Surgeon in Cochin and mother was a highly educated Syrian Christian lady. He graduated in Physics from Loyola College, Madras in 1940 and then did B.E. (Mech) from the University of Madras. He went to USA on a government scholarship to earn his Masters in Mechanical Engineering from Michigan State University from where he passed with distinction.

Dr.Kurien returned from America in 1948 and joined the Dairy Department of Government of India. In May 1949, he was posted as Dairy Engineer at the Government Research Creamery, a small milk-powder factory in Anand. At this time the newly formed cooperative dairy, Kaira District Cooperative MilkProducers Union Limited (KDCMPUL) at Anand, was facing stiff challenges from the privately owned Polson Dairy, which had hitherto monopolized the dairy trade. KDCMPUL had been formed in 1946 under the initiative of SardarVallabhbhai Patel.

Enthused by the challenge of doing something for the milk farmers, Dr.VergheseKurien resigned from his government job and volunteered to help ShriTribhuvandas Patel, the Chairman of KDCMPUL. As Manager, Dr.Kurien provided the necessary administrative and scientific direction to the dairy union leading to its rapid expansion. By 1962, the Cooperative which was organized in 1948 by combining two village milk producer societies stood transformed. Dr.Kurien’s efforts had brought into the Milk Producers’ Union 219 farmer societies with 46,400 members.

The Taste of India…

Dr.Kurien realized the importance of brand in order to market the milk products of the Union. The catchy brand name AMUL (an acronym for Anand Milk Union Ltd.) was derived from the Sanskrit word "Amoolya" which means priceless. The brand which was launched in Oct 1955 and registered in 1957 soon became a household name.

The birth and growth of AMUL tells the oft-documented success story which has served as a model for not just Indian cooperatives but has been replicated in several parts of the world. Amul formed several co-operative societies for a group of villages. These co-operative societies collected milk from the village farmers twice a day. The payment was made to farmers according to the fat content in the milk. Sufficient steps (such as standard fat measurement machine, surprise checks, educating farmers etc.) were taken to prevent malpractices and enhance the overall process. The milk was then transferred to nearby Milk Chiller Unit on the same day.

It was kept in storage for a few hours before being transferred for pasteurization and finally to the cooling and packaging unit. The efficient upstream supply chain was entirely designed by Dr.Kurien and Mr.Tribhuvandas Patel and by the end of 1960 Amul had become a major success story in Gujarat. This cooperative model not only made farmers earn more, but also brought about a social revolution by empowering women, increasing hygiene and medical standards of both people and cattle, and breaking caste and religious distinctions.

A singular achievement of Dr.Kurien, a feat at that time thought quite preposterous and impossible by global standards, was being able to produce milk powder from buffalo milk.

National Dairy Development Board (NDDB)

In 1965, the then Prime Minister LalBahadurShastri, created the National Dairy Development Board (NDDB) with Dr.VergheseKurien at the helm to replicate the success story of Amul throughout the country. Dr Kurien was the Chairman of NDDB from 1965 to 1998. Under his leadership the mission achieved thrust and direction with the launch of 'Operation Flood'.

Operation Flood

'Operation Flood' was aimed at replicating and scaling up the Anand Pattern Cooperative model all over India. Dr.Kurien ensured that NDDB set up for the purpose remained autonomous — legally and financially. He insisted on making farmers' cooperatives an integral part of 'Operation Flood'. Started in 1970 and culminating in 1996, it was one of the world's largest rural development programmes.

A National Milk Grid linked milk producers throughout India with consumers in over 700 towns and cities, reducing seasonal and regional price variations while ensuring that the producer gets fair market prices in a transparent manner and on a regular basis. Dr.Kurien as the main architect of "Operation Flood," helped usher a "White Revolution" in India. Popularly known as the Anand Model, the programme assured a better economic future and sustenance for nine million farm households in 75,000 villages in India. NDDB has since integrated 96,000 dairy cooperatives in a milk grid following the 'Anand Pattern'.

Amul and the Anand Pattern

India's White Revolution had its origin in a single small enterprise started in Gujarat State. Amul (Anand Milk Union Limited), the first modern dairy, was successful due to its robust supply chain, hierarchical network of cooperatives, small and large suppliers are connected. Amul is derived from the Sanskrit word 'amulya', which means "priceless". It denoted and symbolized the pride of *Swadeshi* production. In 1946, at the suggestion of SardarVallabhbhai Patel, the farmers in

Kairadistrict formed a cooperative union to supply milk directly to the Bombay Milk Scheme (BMS), cutting out private dairy and middlemen who were then supplying to BMS. The Karia Union began with two societies and a daily milk collection of not more than 200 litres of milk, under the chairmanship of ShriTribhuvandas Patel. Right from the inception of the dairy cooperative itself, a vital link was established betweenthe producer and Bombay'smarket, ensuring the incentive of astable and remunerative price to thefarmer.

It was particularly fortunate to have two forceful personalities, T.K. Patel and V. Kurien, who were the driving force behind the expansion of output and the introduction of modern methods of processing and marketing. V. Kurien is known as the 'father of operations flood'. The essence of the cooperative's success, however, has been the fact that it is owned by the fanners themselves. They feel that it is their organisation and that they have a stake in its future. Although they employ professionals to run the system, producers can participate in decision-making in various ways. A visit by the Prime Minister, LalBahadurShastri, to the scheme in 1964, convinced him that this was a type of cooperative that worked. The government of India over the years has attempted to stimulate cooperative effort in the countryside, but their good intentions have been frustrated in most states by the vested interests of the local village elites, who have selfishly creamed off the benefits for themselves. The KDCMPU seemed to Mr. Shastri an exception to this rule. He sought to have the structure of the Kaira cooperative reproduced in other parts of India, so that as many farmers as possible could share a new prosperity.

Plate:LalBahadurShastri with the Milkmen

GCMMF

In 1973, Dr.Kurien set up GCMMF (Gujarat Cooperative Milk Marketing Federation) which brought under its umbrella the various village cooperatives of Gujarat. They could now compete with the private sector as a combined and stronger force. From 1973 to 2005 Dr.Kurien was the Chairman cum Managing Director of GCMMF. Under his stewardship India became the largest producer of milk in the world. Today, GCMMF is a 18,000 crore federation and Amul is one of India's strongest names. With the demise of Dr.Kurien, the song might have ceased, but the melody lingers on. At present, his model of cooperative dairy development links over 10 million farmers at 200 dairies across India producing over 20 million litres of milk every day.

TABLE 6.1: MILK: NATIONAL PRODUCTION AND OPERATION FLOOD PROCUREMENT AND SALES

	MILK PRODUCTION (*Million Metric Tons)*	OPERATION FLOOD MILK SALES	OPERATION FLOOD MILK PROCUREMENT	PERCENTAGE
1951	17.41			
1956	19.72			
1961	20.38			
1966	19.37			
1969/70	20.74			
1970/71		0.37	0.14	
1971/72	22.50	0.39	0.41	1.8
1975/76		0.58	0.43	
1977/78	28.40	0.81		
1980/81		1.05	0.96	
1981/82	34.30	1.22		
1985/86	44.00	2.23	2.96	6.7
1988/89		2.65	2.99	
1990/91	54.90	3.02		
1991/92		3.16	3.52	
1992/93		3.16	3.98	
1993/94	60.80	3.23	4.17	6.9
1994/95	63.31[a]	3.52	4.05	6.4
1995/96	66.31[b]			

a. Dairy India 1997, p. 160, "anticipated."
b. Dairy India 1997, p. 160, "target."
Notes:
1. The production estimates are total production and are based on a sample survey. Purchases and sales are from NDDB.
2. For some reason the production survey does not include the amount marketed, so there are no official estimates of total marketings.
3. To some extent this compares different things, since the fat content of milk can vary from say 3.0 percent in a poor Friesian to 12 percent in a good buffalo. A proper accounting would include data on butterfat and solids-not-fat (SNF) produced.
Source: Aneja 1994; NDDB undated; and Gupta 1997, Table 108, and study data.

Source: Candler and Kumar, 1998: 34

White Revolution and Its Impacts on Indian Economy

India is a vast South Asian country with diverse terrain and a huge population. It is suggested by the experts that the population is growing faster than other populous nations. As per the experts, the population boom could pose certain challenges and have a major impact on India's economy. Milk production is one of the major challenges that India had faced before Project Flood.

In India, milk is obtained from animals such as cows, buffaloes, goats and from camels (in Rajasthan). The Indian Dairy Industry has gained substantial expansion from the Eighth Plan onwards. It is believed that India’s milk output has placed the industry first in the world. Additionally, India’s milk output has represented continued growth in the availability of milk and milk products.

As per the Economic data, dairy sector is now the largest contributor in the agricultural sector to the nation’s GDP. India’s dairy industry achieved this success after the occurrence of Project Flood known as White Revolution.

In 1965, the National Dairy Development Board (NDDB) was placed to support, plan and organize dairy development through cooperatives. These cooperatives were envisioned as democratic institutions, owned and managed by rural producers. The project was divided into 3 different phases that helped the National Dairy Development Board to achieve the set targets.

It may sound surprising but it is a fact that India is the largest milk producing country in the world. As per the latest data, the country produced more than 150 million tons of milk in 2017-18 which accounts for around about 17 per cent of total milk production in the world. Milk producing capacity of Indian cows and she-buffaloes is less but their sheer number puts India right there on the top. Andhra Pradesh, Haryana, Rajasthan and Punjab are major milk producing states of India. India has become the leading producer of milk in the world. The milk production that was about 17 million tons in 1950-51 increased upto 105 million tons in 2007-08. The production of milk has gone up by more than six times when compared with that of the Pre-Independence situation. And this has supported the Indian economy.

Intensive Cattle Development Program (ICDP) was introduced during the 1964-65 for promotion of White Revolution. Later on another program ‘Operation Flood’ was introduced by National Dairy Development Board to provide further impetus to White revolution.

Dr.VergheseKurien is called the father of white revolution in India. The founder-chairman of the National Dairy Development Board (NDDB) and the Gujarat Cooperative Milk Marketing Federation ushered India’s White Revolution and was the driving force behind India’s dramatic transformation to world's largest milk producer from a milk-deficient nation. The resounding success of Operation Flood doubled milk available per person and increased milk output four-fold within three decades. Operation Flood made dairy farming India's largest self-sustaining industry and the largest rural employment provider. It contributed mightily to India’s economic as well as social growth as the small farmers accounted for about 70 per cent of milk production, and 22.5 per cent of the rural households' income came from milk.

VergheseKurien was known as the "Father of the White Revolution" for his Operation Flood; it was his idea and thought which transformed India with a milk-deficient nation to the world's largest milk producer. In 1998 India left United States and increased its production. Due to Kurien efforts dairy farming became India's largest self-sustaining industry. It was him who founded around 30 institutions (like AMUL, GCMMF, IRMA, NDDB) which are owned, managed by farmers and run by professionals providing them earning wages. Moreover he is the founder chairman of the Gujarat Co-operative Milk Marketing Federation (GCMMF). He was responsible for the creation and success of the Amul brand of dairy products. Which mad India world’s biggest producer and consumer of dairy ondate.

The small and poor farmers and the landless labourers are benefitted from the White Revolution as they are getting employed and helping to raise and feeding their family. The White Revolution made a positive impact on rural areas and encouraged them to take dairying as an occupation.

Moreover for the improvement in the quality of livestock, extensive cross breeding has been launched. And for a good and positive response of the white revolution, research centres are set up at Anand, Mehsana, and Palanpur (Banaskantha). Moreover, there are three regional centres that are running in Siliguri, Jalandhar, and Erode. Currently there are metro dairies in 10 metropolitan cities of the country, beside 40 plants which are handling more than one lakh litres of milk.

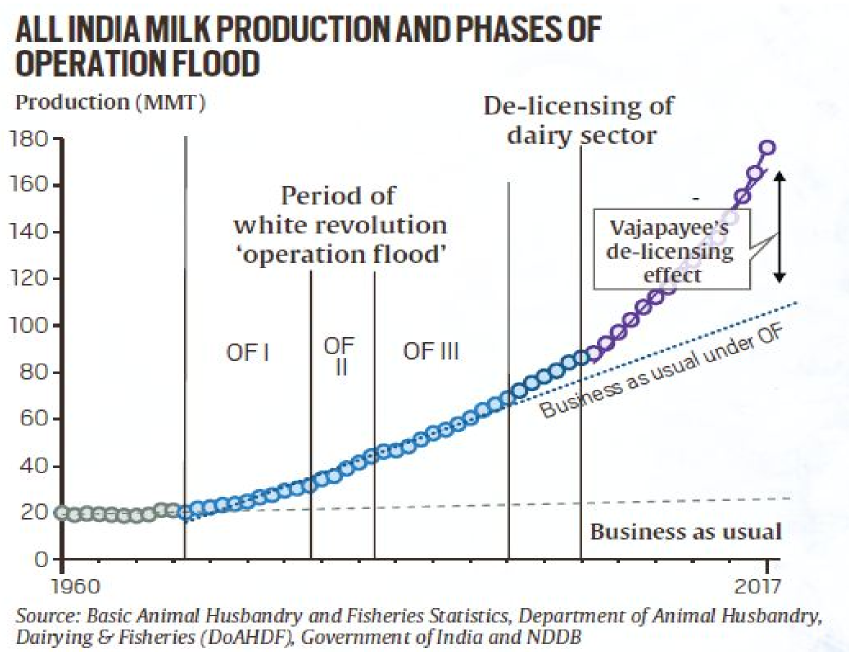


Figure: Effects of White Revolution

Unit-3

2.6 Concept of Industrial Region and Industrial Complex

Secondary activities or manufacturing change raw materials into products of more value to people. As you have seen pulp was changed into paper and paper into a note book. These represent the two stages of the manufacturing process.

The paper made from pulp and cloth made from cotton have had value added to them at each stage of the manufacturing process. In this way the finished product has more value and utility than the raw material that it is made from.

Industry refers to an economic activity that is concerned with production of goods, extraction of minerals or the provision of services. Thus we have iron and steel industry (production of goods), coal mining industry (extraction of coal) and tourism industry (service provider). Classification of Industries Industries can be classified on the basis of raw materials, size and ownership.

Raw Materials: Industries may be agro based, mineral based, marine based and forest based depending on the type of raw materials they use. **Agro based industries** use plant and animal based products as their raw materials. Food processing, vegetable oil, cotton textile, dairy products and leather industries are examples of agro-based industries. **Mineral based industries** are primary industries that use mineral ores as their raw materials. The products of these industries feed other industries. Iron made from iron ore is the product of mineral based industry. This is used as raw material for the manufacture of a number of other products, such as heavy machinery, building materials and railway coaches. **Marine based industries** use products from the sea and oceans as raw materials. Industries processing sea food or manufacturing fish oil are some examples. **Forest based industries** utilise forest produce as raw materials. The industries associated with forests are pulp and paper, pharmaceuticals, furniture and buildings.

Size: It refers to the amount of capital invested, number of people employed and the volume of production. Based on size, industries can be classified into **small scale** and **large scale** industries. Cottage or household industries are a type of small scale industry where the products are manufactured by hand, by the artisans. Basket weaving, pottery and other handicrafts are examples of cottage industry. **Small scale industries** use lesser amount of capital and technology as compared to large scale industries that produce large volumes of products. Investment of capital is higher and the technology used is superior in large scale industries. Silk weaving and food processing industries are small scale industries(Fig 5.1). Production of automobiles and heavy machinery are large scale industries.

Ownership: Industries can be classified into private sector, state owned or public sector, joint sector and cooperative sector. **Private sector industries** are owned and operated by individuals or a group of individuals. **The public sector industries** are owned and operated by the government, such as Hindustan Aeronautics Limited and Steel Authority of India Limited. Joint sector industries are owned and operated by the state and individuals or a group of individuals. MarutiUdyog Limited is an example of joint sector industry. Co-operative sector industries are owned and operated by the producers or suppliers of raw materials, workers or both. Anand Milk Union Limited and Sudha Dairy are a success stories of a co-operative venture.

Factors Affecting Location of Industries

The factors affecting the location of industries are the availability of raw material, land, water, labour, power, capital, transport and market. Industries are situate d where some or all of these factors are easily available. Sometimes, the government provides incentives like subsidised power, lower transport cost and other infrastructure so that industries may be located in backward areas. Industrialisation often leads to development and growth of towns and cities.

Industrial System

An industrial system consists of inputs, processes and outputs. The inputs are the raw materials, labour and costs of land, transport, power and other infrastructure. The processes include a wide range of activities that convert the raw material into finished products. The outputs are the end product and the income earned from it. In case of the textile industry the inputs may be cotton, human labour, factory and transport cost. The processes include ginning, spinning, weaving, dyeing and printing. The output is the shirt you wear.

Industrial Regions

Industrial regions emerge when a number of industries locate close to each other and share the benefits of their closeness. Major industrial regions of the world are eastern North America, western and central Europe, Eastern Europe and eastern Asia. Major industrial regions tend to be located in the temperate areas, near sea ports and especially near coal fields. India has several industrial regions like Mumbai-Pune cluster, Bangalore-Tamil Nadu region, Hugli region, Ahmedabad-Baroda region, Chottanagpur industrial belt, Vishakhapatnam-Guntur belt, Gurgaon-Delhi-Meerut region and the Kollam-Thiruvanathapuram industrial cluster.

Distribution of Major Industries

The world's major industries are the iron and steel industry, the textile industry and the information technology industry. The iron and steel and textile industry are the older industries while information technology is an emerging industry. The countries in which iron and steel industry is located are Germany, USA, China, Japan and Russia. Textile industry is concentrated in India, Hong Kong, South Korea, Japan and Taiwan. The major hubs of Information technology industry are the Silicon valley of Central California and the Bangalore region of India.

Iron and Steel Industry

Like other industries iron and steel industry too comprises various inputs, processes and outputs. This is a feeder industry whose products are used as raw material for other industries. The inputs for the industry include raw materials such as iron ore, coal and limestone, along with labour, capital, site and other infrastructure. The process of converting iron ore into steel involves many stages. The raw material is put in the blast furnace where it undergoes smelting. It is then refined. The output obtained is steel which may be used by other industries as raw material.

Steel is tough and it can easily be shaped, cut, or made into wire. Special alloys of steel can be made by adding small amounts of other metals such as aluminium, nickel, and copper. Alloys give steel unusual hardness, toughness, or ability to resist rust. Steel is often called the backbone of modern industry.

Almost everything we use is either made of iron or steel or has been made with tools and machinery of these metals. Ships, trains, trucks, and autos are made largely of steel. Even the safety pins and the needles you use are made from steel. Oil wells are drilled with steel machinery. Steel pipelines transport oil. Minerals are mined with steel equipment. Farm machines are mostly steel. Large buildings have steel framework.

Before 1800 A.D. iron and steel industry was located where raw materials, power supply and running water were easily available. Later the ideal location for the industry was near coal fields and close to canals and railways. After 1950, iron and steel industry began to be located on large areas of flat land near sea ports. This is because by this time steel works had become very large and iron ore had to be imported from overseas.

In India, iron and steel industry has developed taking advantage of raw materials, cheap labour, transport and market. All the important steel producing centres such as Bhilai, Durgapur, Burnpur, Jamshedpur, Rourkela, Bokaro are situated in a region that spreads over four states — West Bengal, Jharkhand, Odisha and Chhattisgarh. Bhadravati and Vijay Nagar in Karnataka, Vishakhapatnam in Andhra Pradesh, Salem in Tamil Nadu are other important steel centres utilising local resources.

Jamshedpur

Before 1947, there was only one iron and steel plant in the country – Tata Iron and Steel Company Limited (TISCO). It was privately owned. After Independence, the government took the initiative and set up several iron and steel plants. TISCO was started in 1907 at Sakchi, near the confluence of the rivers Subarnarekha and Kharkai in Jharkhand. Later on Sakchi was renamed as Jamshedpur. Geographically, Jamshedpur is the most conveniently situated iron and steel centre in the country.

Sakchi was chosen to set up the steel plant for several reasons. This place was only 32 km away from Kalimati station on the Bengal-Nagpur railway line. It was close to the iron ore, coal and manganese deposits as well as to Kolkata, which provided a large market. TISCO, gets coal from Jharia coalfields, and iron ore, limestone, dolomite and manganese from Odisha and Chhattisgarh. The Kharkai and Subarnarekha rivers ensured sufficient water supply. Government initiatives provided adequate capital for its later development.

In Jamshedpur, several other industrial plants were set up after TISCO. They produce chemicals, locomotive parts, agricultural equipment, machinery, tinplate, cable and wire.

The development of the iron and steel industry opened the doors to rapid industrial development in India. Almost all sectors of the Indian industry depend heavily on the iron and steel industry for their basic infrastructure. The Indian iron and steel industry consists of large integrated steel plants as well as mini- steel mills. It also includes secondary producers, rolling mills and ancillary industries.

Pittsburgh

It is an important steel city of the United States of America. The steel industry at Pittsburgh enjoys locational advantages. Some of the raw material such as coal is available locally, while the iron ore comes from the iron mines at Minnesota, about 1500 km from Pittsburgh. Between these mines and Pittsburgh is one of the world's best routes for shipping ore cheaply – the famous Great Lakes waterway. Trains carry the ore from the Great Lakes to the Pittsburgh area. The Ohio, the Monogahela and Allegheny rivers provide adequate water supply.

Today, very few of the large steel mills are in Pittsburgh itself. They are located in the valleys of the Monogahela and Allegheny rivers above Pittsburgh and along the Ohio River below it. Finished steel is transported to the market by both land and water routes. The Pittsburgh area has many factories other than steel mills. These use steel as their raw material to make many different products such as railroad equipment, heavy machinery and rails.

Cotton Textile Industry

Weaving cloth from yarn is an ancient art. Cotton, wool, silk, jute, flax have been used for making cloth. The textile industry can be divided on the basis of raw materials used in them. Fibres are the raw material of textile industry. Fibres can be natural or man-made. Natural fibres are obtained from wool, silk, cotton, linen and jute. Human made fibres include nylon, polyester, acrylic and rayon.

The cotton textile industry is one of the oldest industries in the world. Till the industrial revolution in the 18th century, cotton cloth was made using hand spinning techniques (wheels) and looms. In 18th century power looms facilitated the development of cotton textile industry, first in Britain and later in other parts of the world. Today India, China, Japan and the USA are important producers of cotton textiles.

India has a glorious tradition of producing excellent quality cotton textiles. Before the British rule, Indian hand spun and hand woven cloth already had a wide market. The Muslins of Dhaka, Chintzes of Masulipatnam, Calicos of Calicut and Gold-wrought cotton of Burhanpur, Surat and Vadodara were known worldwide for their quality and design. But the production of hand woven cotton textile was expensive and time consuming. Hence, traditional cotton textile industry could not face the competition from the new textile mills of the West, which produced cheap and good quality fabrics through mechanized industrial units.

The first successful mechanized textile mill was established in Mumbai in 1854. The warm, moist climate, a port for importing machinery, availability of raw material and skilled labour resulted in rapid expansion of the industry in the region.

Initially this industry flourished in the states of Maharashtra and Gujarat because of favourable humid climate. But today, humidity can be created artificially, and raw cotton is a pure and not weight losing raw material, so this industry has spread to other parts of India. Coimbatore, Kanpur, Chennai, Ahmedabad, Mumbai, Kolkata, Ludhiana, Puducherry and Panipat are some of the other important centres.

Ahmedabad

It is located in Gujarat on the banks of the Sabarmati river. The first mill was established in 1859. It soon became the second largest textile city of India, after Mumbai. Ahmedabad was therefore often referred to as the 'Manchester of India'. Favourable locational factors were responsible for the development of the textile industry in Ahmedabad. Ahmedabad is situated very close to cotton growing area. This ensures easy availability of raw material.

The climate is ideal for spinning and weaving. The flat terrain and easy availability of land is suitable for the establishment of the mills. The densely populated states of Gujarat and Maharashtra provide both skilled and semi-skilled labour. Well-developed road and railway network permits easy transportation of textiles to different parts of the country, thus providing easy access to the market. Mumbai port nearby facilitates import of machinery and export of cotton textiles.

But in the recent years, Ahmedabad textile mills have been having some problems. Several textile mills have closed down. This is primarily due to the emergence of new textile centres in the country as well as non-upgradation of machines and technology in the mills of Ahmedabad.

Osaka

It is an important textile centre of Japan, also known as the 'Manchester of Japan'. The textile industry developed in Osaka due to several geographical factors. The extensive plain around Osaka ensured that land was easily available for the growth of cotton mills. Warm humid climate is well suited to spinning and weaving.

The river Yodo provides sufficient water for the mills. Labour is easily available. Location of port facilitates import of raw cotton and for exporting textiles. The textile industry at Osaka depends completely upon imported raw materials. Cotton is imported from Egypt, India, China and USA. The finished product is mostly exported and has a good market due to good quality and low price. Though it is one of the important textile cities in the country, of late, the cotton textile industry of Osaka has been replaced by other industries, such as iron and steel, machinery, shipbuilding, automobiles, electrical equipment and cement.

Information Technology (IT)

Imagine how much could be accomplished if companies could operate on a twenty-four hour workday. Some software companies in the United States of America and in Bengaluru, India have joined hands to achieve this. There are many ways in which this form of shift work across oceans. For example, two software professionals, Danny in Silicon Valley, California and Smitha in Bengaluru are working on a joint project.

While Smitha in Bengaluru sleeps, Danny in California is working. At the end of his workday, he sends a message to Smitha, updating his progress. When she arrives at work in Bengaluru, a couple of hours later, she notices that a message awaits her. She gets to work on the project straight away. By the end of her workday she relays the results of her efforts back to California. By the way they communicate and work together; it is as if they were sitting in adjoining offices.

The information technology industry deals in the storage, processing and distribution of information. Today, this industry has become global. This is due to a series of technological, political, and socio-economic events. The main factors guiding the location of these industries are resource availability, cost and infrastructure. The major hubs of the IT industry are the Silicon Valley, California and Bengaluru, India. Bengaluru is located on the Deccan Plateau from where it gets the name 'Silicon Plateau'. The city is known for its mild climate throughout the year. Silicon Valley, is a part of Santa Clara Valley, located next to the Rocky Mountains of North America. The area has temperate climate with the temperatures rarely dropping below 0 degrees centigrade. The locational advantages of the Silicon plateau, Bengaluru and Silicon Valley, California are discussed on the next page. You may notice the similarities between the two cities.

There are other emerging information technology hubs in metropolitan centres of India such as Mumbai, New Delhi, Hyderabad and Chennai. Other cities such as Gurgaon, Pune, Thiruvanthapuram, Kochi and Chandigarh are also important centres of the IT industry. However, Bengaluru has always had a unique advantage, as a city with highest availability of middle and top management talent.

World Industrial Regions

- North America
- Europe
- Asia

North America

Manufacturing in North America is concentrated in the north-eastern quadrant of the United States and in south-eastern Canada.

• Only 5 percent of the land area of these countries.., contains one-third of the population and nearly two-thirds of the manufacturing output.

• This manufacturing belt has achieved its dominance through a combination of historical and environmental factors.

Three Determining Location Factors:

•Population Proximity- Early settlement gave eastern cities an advantage to become the country's dominant industrial center.

•Access to Raw Materials- had essential raw materials (by waterway or natural resources).
•Transportation- The Great Lakes and major rivers were an early determining factor is waterway transportation.

•Access to fresh water and power- proximity to great lakes

The major industrial regions of North America are clustered in the northeast U.S. and southeastern Canada, although there are other important centers.

Europe

- **Rhine—Ruhr Valley**

–Transportation- This location at the mouth of Europe's most important river has made Rotterdam the world's largest port.

–Proximity to Raw Materials- Iron and steel manufacturing has concentrated in the Rhine—Ruhr Valley because of proximity to large coalfields. This has resulted in major automotive and machinery manufacturing plants.

–Labour Supply- proximity to large cities provides lots of available labourers

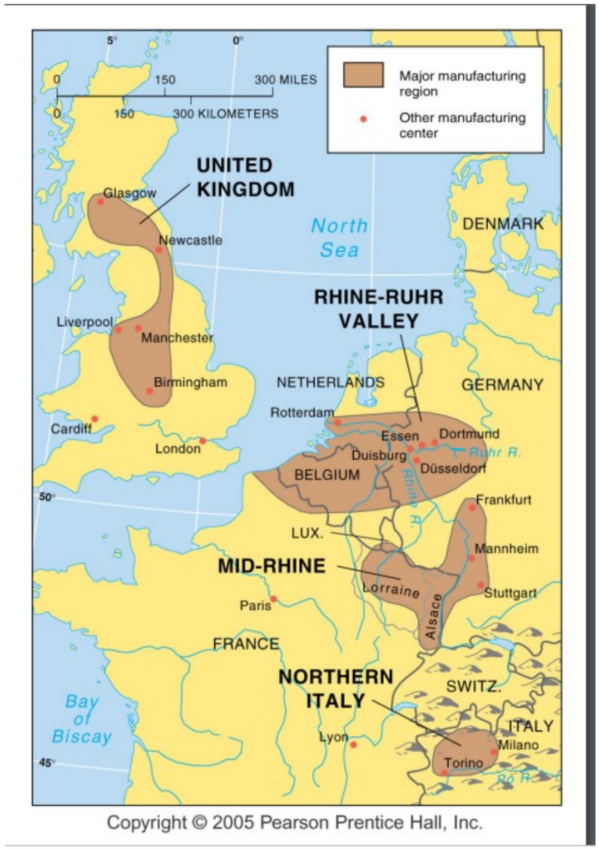


Figure: Rhine-Rhur Valley

Eastern Asia

- Access to Raw Materials- the Asia Pacific region is one of the major growing regions of natural fabrics such as cotton. Also has large coal and iron deposits
- Transportation- Near the oceans and waterways for transport
- Labour Supply- the Asian countries are some of the most densely populated areas of the planet meaning there are a lot of labourers available that also work for low wages

Figure: Industrial Regions of East Asia

Industrial Regions in India

1. Mumbai-Pune Industrial Region
2. Hugli Industrial Region
3. Bangalore-Tamil Nadu Industrial Region
4. Gujarat Industrial Region
5. Chotanagpur Industrial Region
6. Vishakhapatnam-Guntur Industrial Region
7. Gurgaon-Delhi-Meerut Industrial Region
8. Kolfam-Thiruvananthapuram Industrial Region

1. Mumbai-Pune Industrial Region

This region extends from Thane to Pune and in adjoining districts of Nashik and Solapur. In addition, industries have grown at a rapid pace in Kolaba, Ahmednagar, Satara, Sangli and Jalgaon districts also. This region owes its origin to the British rule in India.

The seeds of its growth were sown in 1774 when the island-site was obtained for construction of Mumbai port. The opening of the first railway track of 34 kms between Mumbai and Thane in 1853, opening of the Bhor and Thai Ghats respectively to Pune and Nashik and that of Suez Canal in 1869 led to the development of Mumbai.

The growth of this industrial region is fully connected with the growth of cotton textile industry in India. As the coal was far removed, hydel power was developed in Western Ghats. Cotton was cultivated in the black cotton soil area of the Narmada and Tapi basins.

Cheap labour-force came from the hinterland, the port facilities for export-import and communication links with the peninsular hinterland made Mumbai the ‘Cottonopolis of India’. With the development of cotton textile industry, the chemical industry developed too.

Opening of the Mumbai High petroleum field and erection of nuclear energy plants added additional magnetic force to this region. Now the industrial centres have developed, from Mumbai to Kurla, Kolaba, Thane, Ghatkopar, Ville Parle, Jogeshwari, Andheri, Thane, Bhandup, Kalyan, Pimpri, Pune, Nashik, Manmad, Solapur, Ahmednagar, Satara and Sangli.

In addition to cotton textile and chemical industries, engineering goods, leather, oil refineries; petrochemicals, synthetic and plastic goods, chemicals, drugs, fertilizers, electricals, electronics, software, ship-building, transport and food industries have also developed here.

The partition of the country in 1947 adversely affected this region because 81% of the total irrigated cotton area growing long staple cotton went to Pakistan. Mumbai, the nucleus of this industrial region, is facing the current limitation of space for the expansion of the industry. Dispersal of industries is essential to bring about decongestion.

2. The Hugli Industrial Region

Located in West Bengal, this region extends as a narrow belt running along the river Hugli for a distance of about 100 km from Bansbaria and Naihati in the north to Birlanagar in the south. Industries have also developed in Midnapur district in the west. The river Hugli offered the best site for the development of an inland river port as nucleus for the development of Hugli industrial region.

The old trading centre of late 17th century has developed into the present industrial hub of Kolkata. Thus Kolkata-Haora forms the nucleus of this region. It is very well- connected by the Ganga and its tributaries with the rich hinterland of Ganga-Brahmaputra plains. Besides navigable rivers, roads and the railways provided subsequent links to the great benefit of Kolkata port.

The discovery of coal and iron ore in Chotanagpur plateau, tea plantations in Assam and northern parts of West Bengal and the processing of deltaic Bengal’s jute led to the industrial development in this region. Cheap labour could be found easily from the thickly populated states of Orissa, Bihar, Jharkhand and eastern part of U.P. Kolkata, having been designated capital city of the British India (1773-1912) attracted large scale British investment of capital.

Establishment of first jute mill at Rishra in 1855 ushered in the era of modem industrial clustering in this region. A chain of jute mills and other factories could be established on either side of Hugli River with the help of Damodar valley coal. The port site was best-suited for export of raw materials to England and import of finished goods from that country.

Kolkata’s industries have established by drawing in the raw materials from adjoining regions and distributing the finished goods to consuming points. Thus, the role of transport and communication network has been as important as the favourable locational factors in the growth of this region. By 1921, Kolkata-Hugli region was responsible for two-thirds of factory employment in India.

Just after the partition of old Bengal province in 1947, the region faced, for some years, the problem of shortage of jute as most of the jute-growing areas went to East Pakistan (now Bangladesh). The problem was solved by gradually increasing home production of jute. Cotton textile industry also grew along with jute industry.

Paper, engineering, textile machinery, electrical, chemical, pharmaceuticals, fertilizers and petrochemical industries have also developed in this region. Factory of the Hindustan Motors Limited at Konanagar and diesel engine factory at Chittaranjan are landmarks of this region.

Location of petroleum refinery at Haldia has facilitated the development of a variety of industries. The major centres of this industrial region are Kolkata, Haora, Haldia, Serampur, Rishra, Shibpur, Naihati, Kakinara, Shamnagar, Titagarh, Sodepur, Budge Budge, Birlanagar, Bansbaria, Belgurriah, Triveni, Hugli, Belur, etc.

Alarming rate of silting of the Hugli River was a very serious problem. The depth of water in the channel from bay head to Kolkata docks must be kept at 9.2 metres for big ocean ships to come in. Dredging out of the silt rapidly filling up the water channel was very costly and not a permanent solution to save the life of Kolkata port.

The construction of Farakka barrage about 300 kms upstream on Ganga and flushing of the channel are the only possible answers. The construction of Haldia port in the lower reaches of Hugli to the south of Kolkata is another landmark in relieving the great pressure of cargo ships on the port of Kolkata.

However, the industrial growth of this region has slowed down as compared to the other regions. There are several reasons for this sluggish growth but decline in jute industry is said to be one of the main reasons.

3. Bangalore-Tamil Nadu Industrial Region

Spread in two states of Karnataka and Tamil Nadu, this region experienced the fastest industrial growth in the post-independence era. Till 1960, industries were confined to Bangalore district of Karnataka and Salem and Madurai districts of Tamil Nadu. But now they have spread over all the districts of Tamil Nadu except Viluppuram.

This region is a cotton-growing tract and is dominated by the cotton-textile industry. In fact cotton textile industry was the first to take roots in this region. But it has large number of silk-manufacturing units, sugar mills, leather industry, chemicals, rail wagons, diesel engines, radio, light engineering goods, rubber goods, medicines, aluminium, cement, glass, paper, cigarette, match box and machine tools, etc.

This region is away from the main coal-producing areas of the country but cheap hydroelectric power is available from Mettur, Sivasamudram, Papanasam, Pykara and Sharavati dams. Cheap skilled labour and proximity to vast local market as well as good climate have also favoured the concentration of industries in this region.

Coimbatore has grown rapidly mainly owing to its industrial growth based on Pykara power, local cotton, coffee mills, tanneries, oil presses and cement works. Coimbatore is known as Manchester of Tamilnadu because of its large-scale cotton textile industry. The establishment of public sector units at Banglore like Hindustan Aeronautics, Hindustan Machine Tools, Indian Telephone Industry and Bharat Electronics etc. has further pushed up the growth of industries in the region.

Madurai is known for its cotton textiles. Visvesvarayya Iron and Steel Works is located at Bhadravati. The other important centres of this region are Sivakasi, Tiruchirapalli, Madukottai, Mettur, Mysore and Mandya. Petroleum refinery at Chennai and Narimanam and iron and steel plant at Salem are recent developments.

4. Gujarat Industrial Region

The nucleus of this region lies between Ahmedabad and Vadodara as a result of which it is also known as Ahmedabad-Vadodara industrial region. However, this region extends uptoValsad and Surat in the south and Jamnagar in the west. The region corresponds to the cotton growing tracts of the Gujarat plains and the development of this region is associated with the location of textile industry since 1860s.

This region became important textile region with the decline of cotton textile industry in Mumbai. Mumbai has the disadvantage of paying double freight charges for first bringing the raw cotton from the peninsular hinterland and then despatching the finished products to inland consuming points in India.

But Ahmedabad is nearer the sources of raw material as well as the marketing centres of the Ganga and Satlui plains. Availability of cheap land, cheap skilled labour and other advantages helped the cotton textile industry to develop. This major industrial region of the country, mainly consisting of cotton textile industry, is expanding at a much faster rate in providing a greater factory employment.

The discovery and production of oil at a number of places in the Gulf of Khambhat area led to the establishment of petrochemical industries around Ankleshwar, Vadodara and Jamnagar. Petroleum refineries at Koyali and Jamnagar provide necessary raw materials for the proper growth of petrochemical industries.

The Kandla port, which was developed immediately after independence, provides the basic infrastructure for imports and exports and helps in rapid growth of industries in this region. The region can now boast of diversified industries.

Besides textiles (cotton, silk and synthetic fibres) and petrochemical industries, other industries are heavy and basic chemicals, dyes, pesticides, engineering, diesel engines, textile machinery, pharmaceuticads, dairy products and food processing. The main industrial centres of this region are Ahmedabad, Vadodara, Bharuch, Koyali, Anand, Khera, Surendranagar, Surat, Jamnagar, Rajkot and Valsad. The region may become more important in the years to come.

5. Chotanagpur Industrial Region

As its name indicates, this region is located on the Chotanagpur plateau and extends over Jharkhand, Northern Orissa and Western part of West Bengal. The birth and growth of this region is linked with the discovery of coal in Damodar Valley and iron ore in the Jharkhand-Orissa mineral belt. As both are found in close proximity, the region is known as the 'Ruhr of India'.

Besides raw materials, power is available from the dam sites in the Damodar Valley and the thermal power stations based on the local coal. This region is surrounded by highly populated states of Jharkhand, Bihar, Orissa and West Bengal which provide cheap labour.

The Kolkata region provides a large market for the goods produced in the Chotanagpur region. It also provides the port facility to the region. It has the advantages for developing ferrous metal industries. The Tata Iron and Steel Company at Jamshedpur, Indian Iron Steel Co., at Bumpur-Kulti, Hindustan Steel Limited at Durgapur, Rourkela and Bokaro are the important steel plants located in this region.

Heavy engineering, machine tools, fertilizers, cement, paper, locomotives and heavy electricals are some of the other important industries in this region. Important nodal centres of this region are Ranchi, Dhanbad, Chaibasa, Sindri, Hazaribagh, Jamshedpur, Daltonganj, Garwa and Japla.

6. Vishakhapatnam-Guntur Industrial Region

This industrial region extends from Vishakhapatnam district in the north-eastern part of Andhra Pradesh to Kurnool and Prakasham districts in the south-east and covers most of the coastal Andhra Pradesh. The industrial development of this region mainly depends upon Vishakhapatnam and Machilipatnam ports.

Developed agriculture and rich mineral resources in the hinterlands of these ports provide solid base to the industrial growth in this region. Coal fields of the Godavari basin are the main source of energy. Hindustan Shipyard Ltd. set up at Vishakhapatnam, set up in 1941 is the main focus.

Petroleum refinery at Vishakhapatnam facilitated the growth of several petrochemical industries. Vishakhapatnam has the most modern iron and steel plant which have the distinction of being the only plant in India having coastal location. It uses high quality iron ore from Bailadila in Chhattisgarh.

One lead-zinc smelter is functioning in Guntur district. The other industries of this region include sugar, textiles, paper, fertilizers, cement, aluminium and light engineering. The important industrial centres of this region are Vishakhapatnam, Vijaywada, Vijaynagar, Rajahmundry, Kurnool, Elum and Guntur. Recent discovery of natural gas in Krishna- Godavari basin is likely to provide much needed energy and help in accelerated growth of this industrial region.

7. Gurgaon-Delhi-Meerut Industrial Region

This region developed after independence, but is one of the fastest growing regions of India. It consists of two industrial belts adjoining Delhi. One belt extends over Agra-Mathura-Meerut and Saharanpur in U.P. and the other between Faridabad-Gurgaon- Ambala in Haryana.

The region is located far away from the mineral and power resources, and therefore, the industries are light and market oriented. The region owes its development and growth to hydro-electricity from Bhakra-Nangal complex and thermal power from Harduaganj, Faridabad and Panipat.

Sugar, agricultural implements, vanaspati, textile, glass, chemicals, engineering, paper, electronics and cycle are some of the important industries of this region. Software industry is a recent addition, Agra and its environs have glass industry. Mathura has an oil refinery with its petro-chemical complex. One oil refinery has been set up at Panipat also.

This will go a long way to boost the industrial growth of this region. Gurgaon has Maruti car factory as well as one unit of the IDPL. Faridabad has a number of engineering and electronic industries. Ghaziabad is a large-centre of agro-industries. Saharanpur and Yamunanagar have paper mills. Modinagar, Sonipat, Panipat and Ballabhgarh are other important industrial nodes of this region.

8. Kollam-Thiruvananthapuram Industrial Region

This is comparatively small industrial region and spreads over Thiruvananthapuram, Kollam, Alwaye, Emakulam and Allapuzha districts of south Kerala. The region is located far away from the mineral belt of the country as a result of which the industrial scene here is dominated by agricultural products processing and market oriented light industries.

Plantation agriculture and hydroelectricity provide the industrial base to this region. The main industries are textiles, sugar, rubber, match box, glass, chemical fertilizers, food and fish processing, paper, coconut coir products, aluminium and cement. Oil refinery set up in 1966 at Kochi provides solid base to petrochemical industries. Important industrial centres are Kollam, Thiruvananthapuram, Alluva, Kochi, Alappuzha and Punalur.

Besides the above mentioned eight major industrial regions, India has 13 minor industrial regions and 15 industrial districts. Their names are mentioned below:

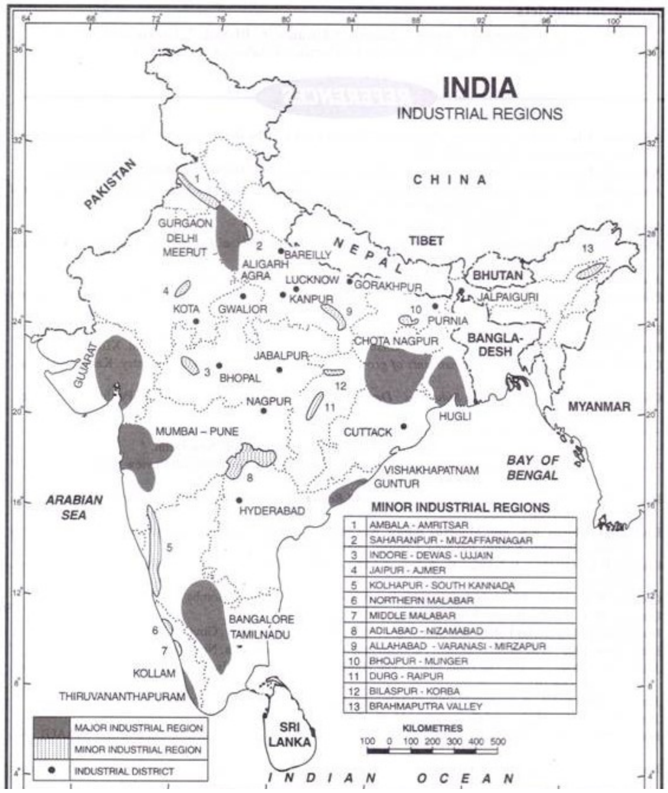


Figure: Industrial Regions of India

Minor Industrial Regions

1. Ambala-Amritsar in Haryana-Punjab.
2. Saharanpur-Muzaffamagar-Bijnaur in Uttar Pradesh.
3. Indore-Dewas-Ujjain in Madhya Pradesh.
4. Jaipur-Ajmer in Rajasthan.
5. Kolhapur-South Kannada in Maharashtra-Karnataka.
6. Northern Malabar in Kerala.
7. Middle Malabar in Kerala.
8. Adilabad-Nizamabad in Andhra Pradesh.
9. Allahabad-Varanasi-Mirzapur in Uttar Pradesh.
10. Bhojpur-Munger in Bihar.
11. Durg-Raipur in Chhattisgarh.
12. Bilaspur-Korba in Chhattisgarh.
13. Brahmaputra Valley in Assam.

Industrial Districts

1. Kanpur, 2 Hyderabad, 3. Agra, 4. Nagpur, 5 Gwalior, 6. Bhopal, 7.Lucknow, 8.Jalpaiguri, 9.Cuttack, 10.Gorakhpur, 11.Aligarh, 12.Kota, 13.Pumia, 14.Jabalpur, 15. Bareilly.

2.7 Special Economic Zone (SEZ)

A special economic zone (SEZ) is an area in which business and trade laws are different from the rest of the country. SEZs are located within a country's national borders, and their aims include: increased trade, increased investment, job creation and effective administration. To encourage businesses to set up in the zone, financial policies are introduced. These policies typically regard investing, taxation, trading, quotas, customs and labour regulations. Additionally, companies may be offered tax holidays, where upon establishing in a zone they are granted a period of lower taxation.

The creation of special economic zones by the host country may be motivated by the desire to attract foreign direct investment (FDI). The benefits a company gains by being in a special economic zone may mean that it can produce and trade goods at a lower price, aimed at being globally competitive. In some countries, the zones have been criticized for being little more than labour camps, with workers denied fundamental labour rights.

Definition

The operating definition of a SEZ is to determine individually by each country. According to the World Bank in 2008, the modern day special economic zone typically includes a 'geographically limited area, usually physically secured (fenced-in); single management/administration; eligibility for benefits based upon physical location within the zone; separate customs area (duty-free benefits) and streamlined procedures.' Special economic zones are those industrial zones which have been set up by government of India to attract foreign companies to invest in the country.

Objectives of Special Economic Zones

- Generation of additional economic activity.
- Promotion of exports of goods and services.
- Promotion of investment from domestic and foreign sources.
- Creation of employment.
- Development of infrastructure facilities.

Advantages of SEZs

- 15 year corporate tax holiday on export profit – 100% for initial 5 years, 50% for the next 5 years and up to 50% for the balance 5 years equivalent to profits ploughed back for investment.
- No licence required for import made under SEZ units.
- Duty free import or domestic procurement of goods for setting up of the SEZ units.
- Goods imported/procured locally are duty free and could be utilized over the approval period of 5 years.
- Exemption from customs duty on import of capital goods, raw materials, consumables, spares, etc.
- Exemption from Central Excise duty on the procurement of capital goods, raw materials, and consumable spares, etc. from the domestic market.
- Exemption from payment of Central Sales Tax on the sale or purchase of goods, provided that, the goods are meant for undertaking authorized operations.
- Exemption from payment of Service Tax.
- The sale of goods or merchandise that is manufactured outside the SEZ (i.e, in DTA) and which is purchased by the Unit (situated in the SEZ) is eligible for deduction and such sale would be deemed to be exports.

- The SEZ unit is permitted to realize and repatriate to India the full export value of goods or software within a period of twelve months from the date of export.
- “Write-off” of unrealized export bills is permitted up to an annual limit of 5% of their average annual realization.
- No routine examination by Customs officials of export and import cargo.
- Setting up Off-shore Banking Units (OBU) allowed in SEZs.
- OBU's allowed 100% income tax exemption on profit earned for three years and 50 % for next two years.
- Exemption from requirement of domicile in India for 12 months prior to appointment as Director.
- Since SEZ units are considered as ‘public utility services’, no strikes would be allowed in such companies without giving the employer 6 weeks prior notice in addition to the other conditions mentioned in the Industrial Disputes Act, 1947.
- The Government has exempted SEZ Units from the payment of stamp duty and registration fees on the lease/license of plots.
- External Commercial Borrowings up to $ 500 million a year allowed without any maturity restrictions.
- Enhanced limit of Rs. 2.40 crores per annum allowed for managerial remuneration.

Disadvantages

- Revenue losses because of the various tax exemptions and incentives.
- Many traders are interested in SEZ, so that they can acquire at cheap rates and create a land bank for themselves.
- The number of units applying for setting up EOU's is not commensurate to the number of applications for setting up SEZ's leading to a belief that this project may not match up to expectations.

Example

Falta Special Economic Zone, Rajarhat and Salt Lake Electronic City of West Bengal; Noida SEZ, Uttar Pradesh; Vishakpatnam SEZ, Telangana; Cochin SEZ, Karnataka etc.

Source: https://www.iastoppers.com/

2.8 Growth of IT Industry in India

Information technology in India is an industry consisting of two major components: IT services and business process outsourcing (BPO). The sector has increased its contribution to India's GDP from 1.2% in 1998 to 7.7% in 2017. According to NASSCOM, the sector aggregated revenues of US$160 billion in 2017, with export revenue standing at US$99 billion and domestic revenue at US$48 billion, growing by over 13%. The United States accounts for two-thirds of India's IT services exports.

History

India's IT Services industry was born in Mumbai in 1967 with the establishment of the Tata Group in partnership with Burroughs.[6] The first software export zone, SEEPZ – the precursor to the modern-day IT park – was established in Mumbai in 1973. More than 80 per cent of the country's software exports were from SEEPZ in the 1980s.

The Indian economy underwent major economic reforms in 1991, leading to a new era of globalization and international economic integration, and annual economic growth of over 6% from 1993–2002. The new administration under recently Lt. AtalBihari Vajpayee (who was Prime Minister from 1998–2004) placed the development of Information Technology among its top five priorities and formed the Indian National Task Force on Information Technology and Software Development.

Videsh Sanchar Nigam Limited (VSNL) introduced Gateway Electronic Mail Service in 1991, the 64 kbit/s leased line service in 1992, and commercial Internet access on a visible scale in 1992. Election results were displayed via National Informatics Centre's NICNET.

Major IT Industries

Google, Microsoft, Facebook, Apple, Wipro, Infosys, Oracle, SanDisk, Harman, Dell, Ericsson, Sabre, Goldman Sachs etc.

Employment generation

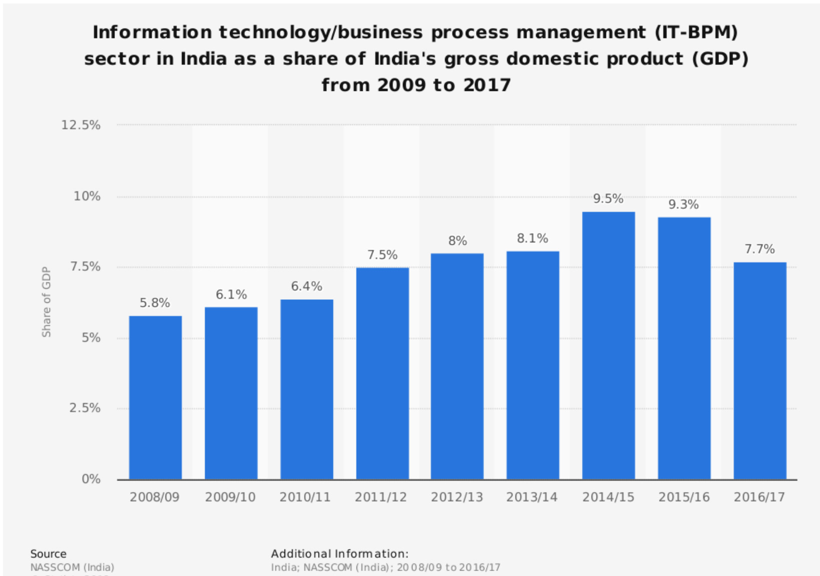
The IT sector has also led to massive employment generation in India. The industry continues to be a net employment generator — expected to add 230,000 jobs in fiscal year 2012, thus directly employing about 2.8 million people and indirectly employing 8.9 million, making it a dominant player in the global outsourcing sector. However, it continues to face challenges of competitiveness in the globalised and modern world, particularly from countries like China and Philippines.

India's growing stature in the Information Age enabled it to form close ties with both the United States and the European Union. However, the recent global financial crises have deeply impacted Indian IT companies as well as global companies. As a result, hiring has dropped sharply, and employees are looking at different sectors like financial services, telecommunications, and manufacturing, which have been growing phenomenally over the last few years.

With fundamental structural changes visible everywhere in the IT services due to Cloud computing, proliferation of Social media, Big data, Analytics all leading to digital services and digital economy, many of the leading companies in India's IT sector reported lower headcounts in their financial results.

Future outlook

The Indian IT market currently focuses on providing low-cost solutions in the services business of global IT. The presence of Indian companies in the product development business of global IT is very meagre, however, this number is slowly on the rise. The other prominent trend is that IT jobs, once confined to Bangalore, are slowly starting to experience a geographical diffusion into other cities like Chennai, Hyderabad and Pune. According to Google estimates, the Indian community of developers will be the largest in the world by 2018.



Source: NASSCOM, India, 2018

Source: NASSCOM, India, 2018

2.9 Industrial Policy

Concept

As a concept of political science, "Policy" has at least two quite distinct meanings. On the one hand "Policies" are often considered to be ways of doing things. In other words they are the decision rules. For example, it is the policy of a particular office to handle the requests in prescribed manner. In this sense policy would answer the questions "how things are being done?" Policy is therefore called administrative policy. Alternatively, policies are often regarded as substantive programmes, referring specifically to how it is being done.

Policies in this sense answer the questions "What is being done and what kinds of problems are handled. The term 'Industrial Policy' refers to the Government's policy towards the establishment of industries, their working and management. It includes all those principals, regulations, rules etc. which would influence the industrialization of the country and also nationalization of industries. It also incorporates the tariff policy, labour policy and government's attitude not only to external assistance but to the public and private sector. Thus it is the plan of the Nation. The present study covers the study of the role of policy analysis in the industry under difference in the industry under difference government's independence.

Need of Industrial Policy of Independent India

When India attained independence, on that time the congress government came into power. Production in India had declined but population was increasing. In view of the need to step up production and to measure inflationary, tendencies, it was necessary to announce an industrial policy which could protect to the some extent economic security. On the other hand inflation was worsened by the economic upheaval of participation of the country and the refugee rehabilitation problem. So, for the growth of industrial structure and thus produce a climate for stimulating investment in Industry the implementation of the industrial policy was necessary. Thus promotion of foreign direct investment forms an integral part of India's overall development.

Earlier, our attitude to foreign investment was also rigid. Foreign technology was permitted if Indian technology was not available, but foreign capital was frowned upon. The changes in policies concerning foreign technology and foreign investment will enable Indian industries to forge ahead with the assistance of foreign investors and suppliers of technology which has not been possible in the past. It is, however, important here to evaluate our economy in a slightly more focused manner from early post independent period to 1991.

The growth performance, though an improvement over the pre-independence period, has been quite lackluster. Therefore, soon after independence, our objectives were directed towards achieving self-reliance, emerging as a strong industrialized nation and, above all, doing this within the framework of social justice.

Main features

Objectives of the Industrial Policy of the Government are –

- ♣to maintain a sustained growth in productivity;
- ♣to enhance gainful employment;
- ♣to achieve optimal utilisation of human resources;
- ♣to attain international competitiveness and
- ♣to transform India into a major partner and player in the global arena.

Policy focus is on –

- ♣ Deregulating Indian industry;
- ♣ Allowing the industry freedom and flexibility in responding to market forces and
- ♣Providing a policy regime that facilitates and fosters growth of Indian industry.

Policy measures

Some of the important policy measures announced and procedural simplifications undertaken to pursue the above objectives are as under:

i) Liberalization of Industrial Licensing Policy

The list of items covered under compulsory licensing under the Industries (Development & Regulation) Act, 1951 is reviewed on an ongoing basis. At present, only five industries are under compulsory licensing mainly on account of environmental, safety and strategic considerations. They are:

1. Distillation and brewing of alcoholic drinks.(Licensing ceased by DIPP in compliance with Supreme Court Order of 29.1.1997 in Bihar Distillery case which ruled that industries engaged in manufacture of potable alcohol would be under the jurisdiction of the States).

2. Cigars and cigarettes of tobacco and manufactured tobacco substitutes.

3. Electronic Aerospace and defense equipment: all types.

4. Industrial explosives including detonating fuses, safety fuses, gun powder, nitrocellulose and matches.

5. Specified Hazardous chemicals i.e. (i) Hydrocyanic acid and its derivatives, (ii) Phosgene and its derivatives and (iii) Isocyanates&diisocyanates of hydrocarbon, not elsewhere specified(example Methyl isocyanate) .

Similarly, from 8 industries reserved for the public sector in 1991, there are only following 2 industries reserved for public sector at present:

1. Atomic energy – Production, separation or enrichment of special fissionable materials and substances and operation of the facilities, specified in DIPP Notification No. S.O.2630 (E) dated 19.10.2009

2. Railway transport.

ii) **Industrial Entrepreneurs' Memorandum (IEM)**

Industries not covered under compulsory licensing are to file an Industrial Entrepreneurs' Memorandum (IEM) with the Secretariat for Industrial Assistance (SIA). No industrial approval is required for such exempted industries. Amendments are also allowed to IEMs filed w.e.f. 1.7.98.

ii) Policy for Small Scale Industries

At present, the SSI is defined under the Micro, Small & Medium Enterprises (MSMED) Act, 2006. Though reservation of items exclusively for the Small Scale sector forms a significant aspect of the industrial policy, review for dereservation of such items is also undertaken by the Government at periodic intervals, in order to enhance competitiveness of such products in the domestic/ global markets. Review of reserved items is thus a continuous process. During the last 5 years itself more than 600 items have been dereserved. At present 20 items are reserved for manufacture in the small scale sector.

All undertakings other than the small scale industrial undertakings engaged in the manufacture of items reserved for manufacture in the small scale sector are required to obtain an industrial licence and undertake an export obligation of 50% of the annual production. This condition of licensing is, however, not applicable to those undertakings operating under 100% Export Oriented Undertakings Scheme, the Export Processing Zone (EPZ) or the Special Economic Zone Schemes (SEZs).

In tune with the provisions of the MSMED Act, 2006 and policy of liberalization, the provision of restricting equity participation by non-SSI undertakings in SSI undertakings upto 24% has been removed by way of rescinding notification No.S.O. 857(E) dated 10.12.1997 vide Notification No.S.O.563(E) dated 27.2.2009.

iii) Non-Resident Indians Scheme

The general policy and facilities for Foreign Direct Investment are applicable to NRIs as well. In addition, Government has extended some additional facilities to NRIs, which include investment in the real estate and civil aviation sectors upto 100 per cent, besides a liberal investment regime on non-repatriation basis.

iv) Electronic Hardware Technology Park (EHTP)/Software Technology Park (STP) Scheme

For building up strong electronics industry and with a view to enhancing export, two schemes viz. Electronic Hardware Technology Park (EHTP) and Software Technology Park (STP) are in operation. Under EHTP/STP scheme, the inputs are allowed to be procured free of duties.

The Directors of STPs have powers to approved fresh STP/EHTP proposals and also grand post-approval amendment in repsect of EHTP/STP projects as have been given to the Development Commissioners of Export Processing Zones in the case of Export Oriented Units. All other application for setting up projects under these schemes, are considered by the Inter-Ministerial Standing Committee (IMSC) Chaired by Secretary (Information Technology). The IMSC is serviced by the SIA.

v) Policy for Foreign Direct Investment (FDI)

Promotion of Foreign Direct Investment (FDI) forms an integral part of the Industrial Policy. FDI helps in accelerating economic growth by means of infusion of capital, technology and modern management practices. Government has put in place a liberal and transparent foreign investment regime, wherein FDI, upto 100%, is allowed, under the automatic route, in most sectors/activities. The FDI policy is announced through issue of Consolidated FDI Policy Circulars.

The Department has also strengthened investment facilitation measures through Foreign Investment Implementation Authority (FIIA).

2.10 Role of Liberalization

Concept of Liberalization

Globalization and privatization have become the buzzwords in the current economic scenario. The concepts of liberalization, globalization and privatization are actually closely related to one another. This LPG phenomenon was first initiated in the Indian Economy in 1990 when the Indian Economy experienced a severe crisis. There was decline in the country's export earnings, national income and industrial output. The government had to seek aid from IMF to resolve it's debt problem. That is when the government decided to introduce the New Industrial Policy (NIP) in 1991 to start liberalizing the Indian economy.

Liberalization means elimination of state control over economic activities. It implies greater autonomy to the business enterprises in decision-making and removal of government interference. It was believed that the market forces of demand and supply would automatically operate to bring about greater efficiency and the economy would recover. This was to be done internally by introducing reforms in the real and financial sectors of the economy and externally by relaxing state control on foreign investments and trade.

With the NIP' 1991 the Indian Government aimed at integrating the country's economy with the world economy, improving the efficiency and productivity of the public sector. For attaining this objective, existing government regulations and restrictions on industry were removed. The major aspects of liberalization in India were;

1. Abolition of licensing: NIP'1991 abolished licensing for most industries except 6 industries of strategic significance. They include alcohol, cigarettes, industrial explosives, defenseproducts ,drugs and pharmaceuticals, hazardous chemicals and certain others reserved for the public sector. This would encourage setting up of new industries and shift focus to productive activities.

2. Liberalization of Foreign Investment: While earlier prior approval was required by foreign companies, now automatic approvals were given for Foreign Direct Investment (FDI) to flow into the country. A list of high-priority and investment-intensive industries were delicensed and could invite up to 100% FDI including sectors such as hotel and tourism, infrastructure, software development .etc. Use of foreign brand name or trade mark was permitted for sale of goods.

3. Relaxation of Locational Restrictions: There was no requirement anymore for obtaining approval from the Central Government for setting up industries anywhere in the country except those specified under compulsory licensing or in cities with population exceeding1 million. Polluting industries were required to be located 25 kms away from the city peripheries if the city population was greater than 1 million.

4. Liberalization of Foreign Technology imports: In projects where imported capital goods are required, automatic license would be given for foreign technology imports up to 2 million US dollars. No permissions would be required for hiring foreign technicians and foreign testing of indigenously developed technologies.

5. Phased Manufacturing Programmes: Under PMP any enterprise had to progressively substitute imported inputs, components with domestically produced inputs under local content policy. However NIP'1991 abolished PMP for all industrial enterprises. Foreign Investment Promotion Board (FIPB) was set up to speed up approval for foreign investment proposals.

6. Public Sector Reforms: Greater autonomy was given to the PSUs (Public Sector Units) through the MOUs (Memorandum of Understanding) restricting interference of the government officials and allowing their managements greater freedom in decision-making.

7. MRTP Act: The Industrial Policy 1991 restructured the Monopolies and Restrictive Trade Practises Act. Regulations relating to concentration of economic power, pre-entry restrictions for setting up new enterprises, expansion of existing businesses, mergers and acquisitions .etc. have been abolished.

Liberalisation in India: 1985 and 1991

Till the mid-1980s India followed a strategy of planned economic development based on import substitution. The 1951 Industrial Development Regulation Act had set out the basic cast and machinery of industrial policy. This involved a comprehensive control over the direction and volume 5 of investment through licenses, a large public sector, and foreign exchange controls. Planned import substitution tilted investment flows initially towards heavy and capital goods industries and later towards chemicals, petroleum and durable consumer goods. It is now universally accepted that this highly regulated and protectionist regime spawned a sluggish and high cost manufacturing system that was also dynamically inefficient (Bhagwati and Desai, 1970; Bhagwati and Srinivasan, 1975; Ahluwalia, 1985).

In 1985, the Rajiv Gandhi administration (1984-1991) crystallised the logic for trimming the regulatory system in an effort to rejuvenate industry. These reforms, collectively termed the New Economic Plan, and characterised as liberalisation by stealth, eased entry and expansion of incumbent firms by removing licence requirements over capacity expansion for many classes of firms: firms with assets below a moderate threshold; those located in "backward" areas; firms in scale-critical industries, and firms that were "modernising". Modernisation was encouraged through relaxing controls on the import of capital equipment and technical know-how.

Licenses were "broadbanded" to allow enterprises to adjust their product mixes more easily to changing market conditions. There was some relaxation of the restrictions on “monopoly houses” if their expansion were in “priority industries”. (Government of India, 1985-86; Srivastava, 1996). While these initiatives increased the freedom of incumbent firms to expand, they were less effective in encouraging entry. We might expect freer play of rivalry among incumbents in the period following these reforms. It is likely that among the incumbents, the smaller benefited more from the removal of many restraints.

The second phase of reforms, which was part of the substantial structural adjustment programme of 1991, was more radical. New industrial policy lifted the rules of investment licensing. Restrictions on expansion by monopoly houses were relaxed, rules of foreign investment relaxed, and sectors reserved for the public sector were thrown open to private sector entry and competition. Procedures for foreign direct investment were simplified and trade tariffs reduced. The maximum import tariff was reduced to 40 per cent from 340 per cent. Quantitative restrictions were eliminated for capital and intermediate goods.

The substantial thrust of 1991 reforms was to expose incumbent firms to greater domestic as well as international competition. Again, one might expect to see an increase in rivalry in the period following these reforms. Larger firms with their greater resources were arguably better placed to respond to the fiercer competition with a range of investments. The relative strengths of the reform episodes of the mid-Eighties and the Nineties and the difference they created in terms of economic growth has been debated. Panagariya (2004) taking issue with 6 DeLong (2004) concludes that the former were "limited in scope and without a clear roadmap", but they laid the basis for the reforms in the 1990s which were "more systematic and systemic".

Impact of Economic Liberalisation

Globalization and liberalization has greatly influenced the Indian economy and made it a huge consumer market. Today, most of the economic changes in the country are based on the demand supply cycle and other economic factors. Today, India is the world"s 12th largest economy in terms of market exchange rate and 4th largest in terms of the Purchasing Power Parity. According to a report by the World Bank, the Indian market is expected to grow at around 8 per cent in the year 20105.

Globalization and liberalization has also made a positive impact on various important economic segments. Today, the service sectors, industrial sectors and the agriculture sector have really grown to a great extent. Around 54 per cent of the annual Gross Domestic Product (GDP) of India comes from the service industry while the industrial and agriculture sector contributes around 29 per cent and 17 per cent respectively.

With the improvement of the market, more and more new sectors are coming up and reaping profits such as IT services, chemical, textiles, cement industry and so on. With the increase in the supply level, the rate of employment is also increasing considerably. There has been an improvement in the manufacturing sector as well which grew from 8.98 per cent in 2005 to around 12 per cent. The communication segment has grown up to around 16.64 per cent. The condition is expected to improve further with more demand and increase in customer base. The yearly growth of the industrial sector has been around 6.8 per cent which will rise more in the future. India is one of the well-known industrial markets in the Asia-Pacific region.

Foreign Direct Investment (FDI):

One of the main aspects of globalization is foreign investment. India today has emerged as one of the perfect markets for foreign investors due to its vast market base. More and more foreign companies are investing in the Indian market to get more returns. The foreign institutional investments (FII) amounts to around US$ 10 billion in FY 2008-09, while the rate of Foreign direct investments (FDI) has grown around 85.1 per cent in 2009 to US$ 46.5 billion from US$ 25.1 billion (2008).

Other Benefits:

Other benefit of economic liberalisation of India since 1991 is under 10;

• Commerce as a percentage of gross world products has increased in 1986 from 15 per cent to nearly 27 per cent in recent years.

• The stock of foreign direct investment resources has increased rapidly as a percentage of gross world products in the past twenty years.

• For the purpose of commerce and pleasure, more and more people are crossing national borders. Globally, on average nations in 1950 witnessed just one overseas visitor for every 100 citizens? By the mid-1980s it increased to six and ever since the number has doubled to 12.

• Worldwide telephone traffic has tripled since 1991. The number of mobile subscribers has elevated from almost zero to 1.8 billion indicating around 30 per cent of the world population. Internet users will quickly touch 1 billion.

2.11 Privatization and Globalization

According to Starr (1988), privatization is a fuzzy concept that evokes sharp political reactions. It covers a great range of ideas and policies, varying from the eminently reasonable to the wildly impractical. Yet however varied and at times unclear in its meaning, privatization has unambiguous political origins and objectives. It emerges from the countermovement against the growth of government in the West and represents the most serious conservative effort of our time to formulate a positive alternative. Privatization proposals do not aim merely to return services to their original location in the private sphere. Some proposals seek to create new kinds of market relations and promise results comparable or superior to conventional public programs. Hence it is a mistake to define and dismiss the movement as simply a replay of traditional opposition to state intervention and expenditure. The current wave of privatization initiatives opens a new chapter in the conflict over the public-private balance.

Privatization as an Idea

In the ideological world we inhabit, contesting interests and parties use "public" and "private" not only to describe but also to celebrate and condemn. Any serious inquiry into the meaning ofprivatization must begin, therefore, by unloading the complexfreight that the public-private distinction carries.

The Public-Private Distinction and the Concept of Privatization

The terms public and private are fundamental to the language of our law, politics, and social life, but they are the source of continual frustration. Many things seem to be public and private at the same time in varying degrees or in different ways. As a result, we quarrel endlessly about whether some act or institution is really one or the other. We qualify the categories: This group is quasi-public, that one is semi-private. In desperation some theorists announce that the distinction is out-dated or so ideologically loaded that it ought to be discarded, or that it is a distinction without a difference.' Yet the terms can hardly be banished nor ought they. To speak intelligently about modern societies and politics without using the words public and private would be as great an achievement as writing a novel with the word "the." However, neither is necessarily the sort of achievement that other theorists or novelists would care to imitate.

The frustration with these ubiquitous categories partly arises because public and private are paired to describe a number of related oppositions in our thought. At the core of many uses are the two ideas that public is too private as open is to closed, and that public is to private as the whole is to the part. In the first sense, we speak of a public place, a public conference, public behaviour, making something public, or publishing an article. The private counterparts, from homes to diaries, are private in that access is restricted and visibility reduced. The concepts of publicity and privacy stand in opposition to each other along this dimension of accessibility. Public is to private as the transparent is to the opaque, as the announced is to the concealed. Similarly, a person's public life is to his or her private life as the outer is to the inner realm.

On the other hand, when we speak of public opinion, public health, or the public interest, we mean the opinion, health, or interest of the whole of the people as opposed to that of a part, whether a class or an individual. Public in this sense often means "common," not necessarily governmental. The public-spirited or public-minded citizen is one concerned about the community as a whole. But in the modern world the concepts of governmental and public have become so closely linked that in some contexts they are interchangeable. The state acts for the whole of a society in international relations and makes rules binding on the whole internally. Public thus often means official. In this sense a "public act" is one that carries official status, even if it is secret and therefore not public in the sense of being openly visible. Indeed, according to the Oxford English Dictionary, private originally signified "not holding public office or official position." As Albert Hirschman points out, this is a meaning that survives in the army "private," that is, the "ordinary soldier without any rank or position." 3 Now, of course, private is contrasted with public to characterize that which lies beyond the state's boundaries, such as the market or the family.

These different contrasts between public and private lead to some apparent conflicts in defining what lies on each side of the boundary. One such conflict concerns the location of the market. To an economist, the marketplace is quintessentially private. But to a sociologist or anthropologist concerned with culture, the marketplace is quintessentially public-a sphere open to utter strangers who nonetheless are able to understand the same rules and gestures in what may be a highly ritualized process of exchange. While economists use the public-private distinction to signify the contrast between state and market, analysts of culture-particularly those concerned with the roles and relations of men and women-take the public sphere to include the market as well as politics and contrast them both with the private domain of the family. In this sense, the public-private distinction is sometimes taken to mark out the contested boundaries of the male and female worlds-a usage that takes us back to the notion of the private as being more closed, more shielded from contact and view, than the open encounters of public life.

From these varying uses of the categories come several contrasting conceptions of the public sphere. The public sphere may be conceived of as the open and visible-the sphere of public life, public theatre, the public marketplace, public sociability. The public sphere also may be conceived of as that which applies to the whole people or, as we say, the general public or the public at large, in which case the public may consist of an aggregate or a mass who have no direct contact or social relation-the very opposite of a sphere of sociability. Or the public sphere may be conceived specifically as the domain circumscribed by the state, although exactly where to draw the state's boundaries may be difficult indeed.

The general meanings of privatization, then, correspond to withdrawals from any of these variously conceived public spheres. Historians and sociologists write about the withdrawal of affective interest and involvement from the sphere of public sociability. For example, in their work on the development of the modern family, Peter Willmott and Michael Young argue that as the modern household became equipped with larger homes, private cars, televisions, and other resources, more time and capital came to be invested in the private interior of the family and less in public taverns, squares, and streets.

Similarly, Richard Sennett suggests that since the eighteenth century modern society has seen a decline of public culture and sociability, a deadening of public life and public space, a privatization of emotion. 6 Such arguments shade into a second meaning of privatization: a shift of individual involvements from the whole to the part-that is, from public action to private concerns-the kind of privatization that Hirschman describes as one swing in a public-private cycle of individual action. In this sort of public-to-private transition, the swing is not from sociability to intimacy but from civic concern to the pursuit of self-interest. Privatization can also signify another kind of withdrawal from the whole to the part: an appropriation by an individual or a particular group of some good formerly available to the entire public or community. Like the withdrawal of involvement, privatization in the sense of private appropriation has obvious implications for the distribution of welfare (Starr. 1988).

From these meanings it is but a short step to the sense of privatization as a withdrawal from the state, not of individual involvements, but of assets, functions, indeed entire institutions. Public policy is concerned with privatization at this level. But the two forms, the privatization of individual involvements and the privatization of social functions and assets, are certainly related, at least by ideological kinship. A confidence that pursuit of private gain servesthe larger social order leads to approval for both self-interested behaviourand private enterprise.

Thus far I have been talking about privatization as if both spheres, public and private, were already constituted. But in a longer perspective, their constitution and separation represent complementary processes. Much historical experience corresponds to Simmel's paradoxical dictum that "what is public becomes ever more public, and what is private becomes ever more private.

This is true specifically of the histories of the state and the family. The difference between patrimonial domination and modern bureaucracies, as Weber describes the two, is precisely that in the patrimonial state public and private roles were mixed and in the modern state these roles are more clearly distinguished." The modern state distinguishes offices and persons.

The office is public, and its files, rules, and financesare distinct from the personal possessions and character of individuals.As public administration and finance were separated from thehousehold and personal wealth of the ruler, the modern state became,in effect more public; the person and family of the ruler, moreprivate.That the domestic sphere has generally become more privateis one of the classic themes of modern sociology and the historyof the family.

The Political Meaning of Privatization

The term privatization did not gain wide circulation in politics until the late 1970s and early 1980s. With the rise of conservative governments in Great Britain, the United States, and France,privatization has come primarily to mean two things: (1) any shift ofactivities or functions from the state to the private sector; and, morespecifically, (2) any shift of the production of goods and services from public to private. Besides directly producing services, governmentsestablish the legal framework of societies and regulate socialand economic life, and they finance services that are privately producedand consumed.

The first, broader definition of privatizationincludes all reductions in the regulatory and spending activity of thestate. The second, more specific definition of privatization excludesderegulation and spending cuts except when they result in a shiftfrom public to private in the production of goods and services. Thismore focused definition is the one that I shall use here. It leavesopen the possibility that privatization may not actually result in lessgovernment spending and regulation-indeed, may even unexpectedlyincrease them.Several further points about my definition need clarification.

First, the public sector here includes agencies administered as partof the state and organizations owned by it, such as state enterprisesand independent public authorities like the British BroadcastingCorporation (BBC) or the Port Authority of New York and NewJersey. In the private sector I include not only commercial firms butalso informal and domestic activities, voluntary associations, cooperatives,and private non-profit corporations.

Second, in the definition I am using, privatization refers to shiftsfrom the public to the private sector, not shifts within sectors. Thusthe conversion of a state agency into an autonomous public authorityor state-owned enterprise is not privatization, though it may wellput the enterprise on a commercial footing.

Third, shifts from publicly to privately produced services may resultnot only from a deliberate government action, such as a sale ofassets, but also from the choices of individuals or firms that a governmentis unwilling or unable to satisfy or control. In many countries,private demand for education, health care, or retirementincome has outstripped public provision. As a result, privateschooling, medical care, and pensions have grown to relativelylarger proportions. This is demand-driven privatization. When privatizationis a demand-driven process, it does not require an absolutereduction in publicly produced services. Stagnation or slow growthin the public sector may be the cause. In some socialist societies thegrowth of an "underground" economy represents a form of privatizationthat is not a planned development (though it may well resultfrom development planning). In other words, as a process, privatizationencompasses more institutional changes than thosebrought about by self-conscious privatization policies. It seems useful,then, to distinguish instances of privatization according towhether they are predominantly policy or demand-driven.

Fourth, if one shifts attention from the sphere of production tothe sphere of consumption, one may alternatively define privatizationas the substitution of private goods for public goods. A publicgood, in the economist's sense, has two distinguishing properties:One person's consumption does not preclude another's; and excludinganyone from consumption is costly, if not impossible. Theprototypical example is fresh air. A public good need not be producedby government. A broadcast television program is a publicgood even if it is provided by a commercially owned station; butvideotape is not, nor is programming on subscription cable services.Any shift toward these forms of non-broadcast television representsa privatization of consumption, even if the local cable service is municipally owned.

Depending on whether one is talking about the locus of productionor the forms of consumption, privatization can mean rather differentthings. In regard to production, "privatization of healthcare" might mean a transfer of medical facilities from public to privateownership; regarding consumption, it might refer to a shift inexpenditures from public health (environmental protection, vaccinations,etc.) to individual medical care. Similarly, "privatizationof transportation" might refer to the conversion of an urban bussystem from public to commercial ownership; or it might mean ashift in ridership from buses to private automobiles, regardless ofwhether the bus company is municipal or commercial. Strictlyspeaking, public transportation is not a public good, since exclusionis possible and only one person at a time can sit in a seat; however,because buses and trains are open to the public at large, commoncarriers are a distinctively public form of consumption compared toprivate cars. More generally, the historical process described byWillmott and Young-the concentration of consumption activitiesin the home-represented a shift toward more privatized forms ofconsumption. This shift has been the source of much criticism ofcontemporary society, as in John Kenneth Galbraith's famous contrastof private opulence and public squalor in The Affluent Society (Starr, 1988).

The Economic Theory of Privatization

Even within the economic theory of privatization, there are some subtle but important differences between two approaches: the radical view of privatization as a reassignment of property rights and the more moderate, conventional view of privatization as an instrument for fine-tuning a three-sector economy.

1. Economic Model 1: Privatization as a Reassignment of Property Rights

Private ownership and competitive markets are normallythought to go hand in hand, but the two issues of ownership andmarket structure are often separate. For the economist devoted toboth, the question then arises as to which object of affection is morebeloved: private ownership or competition. Here a difference ofopinion appears among economists that corresponds to a preferencefor either privatization or liberalization. Those who believethat efficient performance depends on private ownership *per se*favour privatization, even in cases generally regarded as natural monopolies.Conversely, those who see competition as the critical spurto efficiency are more skeptical about the benefits of privatizing monopoliesand often put more emphasis on other policies, such asderegulation. In the case of a government telecommunications monopoly,for example, those who stress ownership may be willing toprivatize the monopoly intact, whereas those who stress competitionmay prefer to break it up before sale or even to keep it in publicownership while allowing private firms to compete with it on equalterms (Starr, 1988).

Thus the perspective that unequivocally points to privatization asdesirable policy holds that property ownership is the fulcrum ofpolitical economy. Curiously, the two unlikely bedfellows sharingthis appreciation of ownership are Marxism and Chicago economics,which draw from it opposite but equally strong conclusions aboutthe overriding importance of getting ownership into the right sector.From the Chicago tradition come two closely related clusters ofwork: the theory of property rights and the theory of public choice.Both attempt to enlarge the conventional economic paradigm bytreating the classical firm and modern package of property rights asonly one of various possible institutional forms.

2. Economic Model 2: Privatization as a Relocation of Economic Functions

Compared to the right-wing schools that condemn the public sector as irredeemably inefficient, policy analysts trained in conventional microeconomics tend to have a more qualified, though still highly critical, view of public institutions. Rather than attribute the performance of public organizations to the incentives created by public ownership per se, mainstream policy analysts generally think of designing the right incentives within the framework of public organization. Of course, the overwhelming consensus is that privateownership is more efficient in providing private goods in competitivemarkets; hence it is rare to find any respectable opinion in favourof government ownership of factories producing high-performancesports cars. Mainstream views do vary, however, about the properrole of public institutions in producing public goods and managingnatural monopolies.

Viewing competition as the critical issue, theneo-classically trained are inclined to favour privatization insofar as itrepresents a move toward competition under conditions when marketsshould be expected to work efficiently. However, in recentyears the requirements for efficient markets have come to be understoodmore liberally, while the reputation of public enterprise hasmarkedly declined. Hence, the prevailing consensus in economicsand policy analysis has become more sympathetic to privatizationthan it was two or three decades ago.

Privatization as Community Empowerment

A different set of arguments, not chiefly concerned with efficiency, comes from a more sociological theory of privatization that emphasizes the strengthening of communities. In the most noteworthy exposition of this position, Peter Berger and Richard Neuhaus propose that government "empower" voluntary associations, community organizations, churches, self-help groups, and other less formal "mediating" institutions that lie between individuals andsociety's "alienating megastructures."In their view, the modernliberal state has undermined these "value-generating," "valuemaintaining," "people-sized institutions" by establishing service bureaucraciesthose take over their functions. Berger and Neuhaus arenot opposed to the provision of social welfare, but they urge that,wherever possible, public policy rely on mediating institutions forthe delivery of publicly financed services.

The view of privatization as community empowerment stands insharp contrast to the conception of privatization as an extension ofproperty rights. Berger and Neuhaus emphatically reject a narrowlyindividualistic view of human motivation. Indeed, they criticize liberalismprecisely for defending individual rights over the rights ofsocial groups to assert their own values; for example, they defendthe capacity of neighbourhoods to sustain "democratically determinedvalues in the public sphere" by exhibiting religious symbolsin public places.They also suggest that attacks on the ideals ofvoluntary service "aid the expansion of the kind of capitalist mentalitythat would put a dollar sign on everything on the grounds thatonly that which has a price tag has worth." Their concern is not toexpand the domain of the profit motive but rather to strengthen local,small-scale forms of social provision. This is privatization with ahuman face, and it bears some resemblance to left-wing interest incommunity organizations and cooperatives (Starr, 1988).

Privatization as a Reduction of Government Overload

A final theory justifying privatization holds that privatization is desirable for its likely political effect in deflecting and reducing demands on the state. In the 1970s, some critics suggested that the Western democracies were suffering from an "overload" of pressure, responsible for excessive spending and poor economic performance. In that framework privatization represents one of several policies encouraging a counterrevolution of declining expectations. In a similar vein, Stuart Butler of the Heritage Foundation has argued that privatization can cure budget deficits by breaking up the kind of public spending coalitions described by public choice theory. Privatizing government enterprises and public services, in this view, will redirect aspirations into the market and encourage a more entrepreneurial consciousness.

The political theory of privatization has several different, overlappingelements. First, the privatization of enterprises is a privatizationof employment relations. The advocates of privatization hopeto divert employees' wage claims from the public treasury, with itsvast capacity for taxing and borrowing, to private employers, whopresumably will have more spine in resisting wage demands. Moreover,the proponents hope for a trickle-down of entrepreneurshipfrom the newly privatized managers to the workers; for that veryreason, privatizers often are perfectly willing to sell to the workers,at an advantageous price, whole enterprises or at least some proportionof the shares. In addition, by shifting to private contractorseven in a few selected areas, government might signal a harder lineon wage concessions and thereby weaken public employee unions.

Second, the advocates of privatization hope also for a privatization of beneficiaries' claims. Instead of marching outside of government offices when things go wrong, the privatizers want them to direct their ire to private service providers-or better yet, simply to switch to other providers. In other words, privatization could mean a wholesale shift, in Hirschman's terms, from "voice" to "exit" as the usual and preferred tactic of coping with dissatisfaction.

Third, the privatization of public assets and enterprises is also aprivatization of wealth. Advocates such as Margaret Thatcher wantprivatization to increase the proportion of the population who ownshares of stock and therefore take a more positive view of profitmaking."People's capitalism" is an old idea, but using privatizationof public assets to bring it about is new. Moreover, by privatizingother assets such as public housing and Social Security trustfunds, privatizers hope to turn public claimants into property ownersand engender in them a deeper identification with capitalism.They expect the worker who receives a retirement income from aprivate pension or individual retirement account to have a moreconservative view of the world than that of the worker who dependson rent subsidies and a government check every month.

This political theory of privatization, like the economicand sociologicaltheories, contains empirical predictions as well as normativejudgments. The predictions concern the probable effects of privatizationon political consciousness and action; the normative judgmentsconcern the desirability of weakening the politicalfoundations of public provision. Empirically, it seems unlikely thatcontracting-out, vouchers, and other arrangements for paying private providers will reduce pressure on government spending; the contractors are as likely as public employees to lobby for larger budgets. However, some forms of privatization may, indeed, change the underlying political values, understandings, and capacities for action in society. Turning public tenants into private homeowners, public employees into private employees, and Social Security beneficiaries into investors in private retirement accounts could very well change their frame of social and political thought. These aspects raise rather different issues from the usual efficiency-minded discussions of privatization; they demand that we consider the meaning of privatization not only as a theory but also as a political practice.

Privatization as a Political Practice

A. The Political Contexts and Uses of Privatization

The meaning of privatization depends in practice on a nation's position in the world economy. In the wealthier countries it is easy to treat privatization purely as a question of domestic policy. But where the likely buyers are foreign, as in the Third World, privatization of state-owned enterprises often means denationalization-a transfer of control to foreign investors or managers. Since state ownership often originally came about in an act of national self-assertion, privatization appears to be a retreat in the face of international pressure. In that sense, national memory colours the meaning of privatization.

However, even in the United States, privatization would be understood rather differently if public assets up for sale or contracts up for bid were likely to be taken over by the Russians or even the Japanese. The more dependent a nation is on foreign investment,the greater the likelihood that privatization will raise theprospect of diminished sovereignty and excite the passions of nationalism. Where privatization raises such issues, it is often blocked,or citizens and domestic firms are reserved exclusive rights to publiclyoffered assets, shares, or contracts. In many Western countries,state ownership owed more in the first place to nationalist than tosocialist sentiment; hence it is scarcely surprising that nationalism isliable to derail or distort privatization plans.

Throughout the world, the privatization of enterprises with strategicmilitary or economic significance raises especially sensitive questionsof sovereignty and security.In most oil-producing countries,for example, no government is likely to try to privatize the state oilcompanies because of the likely domestic political reaction. Even inGreat Britain, the prospective sale of a helicopter company to anAmerican company caused a political stir.

In general, the political uses of privatization are bound to compromisethe avowed efficiency objectives. Governments that are in ahurry to sell state-owned enterprises may make concessions to currentmanagers, whose cooperation is instrumental in divestiture.Privatization then becomes an occasion for managerial enrichmentand entrenchment. It is striking that in Great Britain, France, andother countries that have privatized state-owned enterprises, privatizationusually brings about little or no change in top management (Starr, 1988).

B. Privatization as a Reordering of Claims

Privatization needs to be understood as a fundamental reordering of claims in a society. As it was indicated earlier, in the liberal world the terms public and private sum up a whole structure of rules and expectations about the proper conduct and limits of the state. To say some activity is public is to invoke claims of public purpose, public accountability, and public disclosure. To say something is private is to claim protection from state officials and other citizens (Starr, 1988).

The theory of property rights sees privatization as a reassignment of claims to the control and use of assets, but it misses the special claims of the public sphere in a democratic society-claims for greater disclosure of information, which should improve the social capacity to make choices, and for rights of participation and discussion, which permit the discovery and formation of preferences that are more consistent with long-term societal interests. As a general movement of institutional design, privatization undermines the foundation of claims for public purpose and public services.

Privatization is not only a policy; it is also a signal about the competenceand desirability of public provision. It reinforces the viewthat government cannot be expected to perform well. If, to manyAmericans, private means better, it is partly because of long-existingrestrictions on the scope and quality of public provision. We commonlylimit public services to a functional minimum and therebyguarantee that people will consider the private alternative a step up.This niggardliness shows itself in ways large and small.

Privatization, as some advocates themselves point out, represents an effort to alter the conditions of political competition by breaking up the coalitions supporting public provision and by promoting more market-oriented political values. In other words, it is an attempt to fix in place the conservative orientation that has emerged forcefully in the 1980s. No one needs doubt that public institutions, like private ones, are bases of wealth and power. They are environments that encourage those who work within them to develop different political orientations. To alter the public-private balance is to change the distribution of material and symbolic resources influencing the shape of political life. Privatization ought to be frankly recognized as part of an effort of conservatives to reinforce their own power position.

Privatization of Indian Economy

Indian economy is a major and quickly developing economy – both as far as the offer in the worldwide populace and additionally as far as its creation structure and size. India dwells in excess of one billion populace now and it makes one-6th of the total populace. As far financial advance is concerned, now India positions among a couple of quickest developing economies in the world. We would first be able to take a gander at the example of monetary development in India since 1951. The example of financial development in India has been very inquisitive and its understanding will help us in taking the story forward. Development example of the Indian economy has been since the 1950s when India began taking its own particular autonomous financial choices in the wake of getting freedom from the British run in 1947 (Mittal, 2018).

We can find that the development example of the Indian economy has not been exceptionally smooth. It has been exceptionally uneven as can be seen that there is plainly a gap between a significant lot of the financial development in India where 1980-81 appears to have turned into a genuine defining moment as before that there has been more uneven and moderate monetary execution. In any case, since 1980-81 there has been an upward pattern with better execution quantitatively and additionally subjectively. It can be seen for national salary and additionally for the per capita wage. Private ventures can take quick choices, make opportune speculations, draw in the best ability, and so on. These things can guarantee an opportune enlistment of new innovation, enhance aggressiveness, a wide decision/better cost/esteem/benefit for the shopper and so forth. However, Governments need to guarantee that:

•There is sufficient rivalry in the market
•There is no cartelization or restraining infrastructure
•Proper Regulatory instrument
•Incentivize innovation advancement, generally, individuals search for just here and now benefit expansion.

Major Causes of Privatization are:

- To reduce the burden on Government
- To strengthen competition
- To improve Public finances
- To fund infrastructure growth
- Accountability to shareholders
- To reduce unnecessary interference
- More disciplined labour force.

The private sector has effective policies in solving the problem of externalities, through costless bargaining, driven by individual incentives. According to the Coase Theorem, individual parties will directly or indirectly take part in a cost-benefit analysis, which will eventually result in the most efficient solution.

When comparing with public sector, the private sector responds to incentives in the market. On the other hand, public sector often has non-economic goals. The public sector is not highly driven to maximize production and allocate resources effectively, causing the government to run high cost, low-income enterprises. Privatization directly shifts the focus from political goals to economic goals, which leads to development of the market economy (Poole, 1996). The downscaling aspect of privatization is an important one since bad government policies and government corruption can play a large, negative role in economic growth (Easterly, 2001). Through privatizing, the role of the government in the economy is condensed, thus there is less chance for the government to negatively impact the economy (Poole, 1996).

Privatization may have a positive impact on a country's economic situation. Privatization should not be used to finance new government expenditures and pay off future debts. Instead, privatization enables countries to pay a portion of their existing debt, thus reducing interest rates and raising the level of investment. By reducing the size of the public sector, the government reduces total expenditure and begins collecting taxes on all the businesses that are now privatized. This process can help bring an end to a vicious cycle of over-borrowing and continuous increase of the national debt (Poole, 1996).

Nations around the world have adopted different methods of privatizing state assets depending on the initial conditions of the country's economy and the economic principles of the political party in charge.

Major method of privatization is the sale of state-owned enterprises to private investors. The state would simply decide which institutions should be privatized and through the use of market mechanism, private investors are able to buy shares of each organization. Advantage of this method of privatization is that it creates badly needed revenues for the state while putting privatized firms in the hands of investors who have the incentives and the means of investing and reformation.

Other method of privatization is called voucher privatization. The government universally distributes vouchers to its eligible citizens, which can be sold to other investors or exchanged for shares in other institutions being privatized. Although this method does not create profits for the state, it does privatize state-owned firms in a short period of time.

Next method of privation is called internal privatization, also known as "employee or management buyout". State-owned enterprises are sold to managers (for an extremely low price) who are already familiar with the particular firm and its structure, but there are minimal revenues created for the state. This method creates some incentives but the incentives are much stronger when firms are sold to strategic investors. Furthermore, new owners often do not have the resources to invest and restructure, which is badly needed in a large percentage of state-owned firms in underdeveloped countries (Stirbock, 2001).

One of the noticeable feature of privatization is the improved competitive characteristics it provides to the enterprises which prove to be fruitful for the business as well as the country. Nonetheless, privatization contracts are greatly influenced by merger variables and even global issues and are structured on the basis of manipulation of the government and the private actors along with the administering jurisdiction.

Positive Impact

Privatization may positively affect a nation's monetary circumstance. Privatization ought not to be utilized to fund new government uses and pay off future obligations. Rather, privatization empowers nations to pay a part of their current obligation, in this way decreasing loan fees and raising the level of the venture. By decreasing the measure of general society part, the legislature lessens add up to consumption and starts gathering charges on every one of the organizations that are presently privatized. This procedure can help convey a conclusion to an endless loop of over-acquiring and ceaseless increment of the national obligation.

One of the recognizable elements of privatization is the enhanced aggressive qualities it gives to the ventures which end up being productive for the business and in addition the nation. In any case, privatization contracts are incredibly affected by merger factors and even worldwide issues and are organized based on control of the legislature and the private on-screen characters alongside the regulating ward.

Privatization can be arranged in three sections –

Delegation: Government keeps hold of obligation and private undertaking handles completely or somewhat the conveyance of item and administrations.

Divestment: Government surrenders the obligation.

Displacement: The private venture extends and continuously dislodges the administration element.

Privatization surely is useful for the advance and maintainability of the state-claimed endeavours. The upsides of privatization can be evident from both microeconomic and macroeconomic effects that privatization applies.

Microeconomic focal points

The state claimed undertakings, for the most part, are beaten by the private endeavours intensely. At the point when looked at the last mentioned, it demonstrates better outcomes as far as benefits and effectiveness and profitability. In this manner, privatization can give the vital push to the failure to meet expectations PSUs.

- Privatization realizes major auxiliary changes giving force in the aggressive areas.

- Privatization prompts usage of the worldwide accepted procedures alongside administration and inspiration of the best human ability to encourage feasible upper hand and ad-libbed administration of assets.

Macroeconomic Preferences

Privatization positively affects the monetary development of the division which was beforehand state commanded by a method for diminishing the shortages and obligations.

- The net exchange to the State possessed Enterprises is brought down through privatization.

- It helps in raising the execution benchmarks of the business when all is said in done.

- It can at first undesirably affect the workers yet continuously in the long haul, will demonstrate profitable for the development and thriving of the representatives.

- Privatized endeavours give better and speedy administrations to the customers and help in enhancing the general framework of the nation

Disadvantages

a) Private sector mainly focuses more on profit maximization and less on social objectives dissimilar to public sector that initiates socially viable adjustments in case of emergencies and criticalities.
b) There is lack of clearness in private sector and stakeholders do not get the complete information about the functionality of the enterprise.

c) Privatization has provided the unnecessary support to the corruption and unlawful ways of accomplishments of licenses and business deals amongst the government and private bidders. Lobbying and bribery are the common issues corrupting the practical applicability of privatization.
d) Privatization loses the mission with which the enterprise was established and profit maximization programme encourages malpractices like production of lower quality products, elevating the hidden indirect costs, price escalation etc.
e) Privatization results in high employee turnover and a lot of investment is required to train staff and even making the existing manpower of PSU abreast with the latest business practices.
f) There can be a conflict of interest amongst stakeholders and the management of the buyer private company and initial resistance to change can impede the performance of the enterprise.
g) Privatization intensifies price inflation in general as privatized enterprises do not get government subsidies after the deal and the burden of this inflation affects the common man.

Examples of privatization of companies in India

- Lagan Jute Machinery Company Limited (LJMC)
- Videsh Sanchar Nigam Limited (VSNL)
- Hindustan Zinc Limited (HZL)
- Hotel Corporation Limited of India (HCL)
- Bharat Aluminium Company limited (BALCO)

Globalization of Economy

Globalization essentially means integration of the national economy with the world economy. It implies a free flow of information, ideas, technology, goods and services, capital and even people across different countries and societies. It increases connectivity between different markets in the form of trade, investments and cultural exchanges.

The concept of globalization has been explained by the IMF (International Monetary Fund) as 'the growing economic interdependence of countries worldwide through increasing volume and variety of cross border transactions in goods and services and of international capital flows and also through the more rapid and widespread diffusion of technology.'

The phenomenon of globalization caught momentum in India in 1990s with reforms in all the sectors of the economy. The main elements of globalization were;

1. To open the domestic markets for inflow of foreign goods, India reduced customs duties on imports. The general customs duty on most goods was reduced to only 10% and import licensing has been almost abolished. Tariff barriers have also been slashed significantly to encourage trade volume to rise in keeping with the World trade Organization (WTO) order under General Agreement on Tariff and Trade (GATT).
2. The amount of foreign capital in a country is a good indicator of globalization and growth. The FDI policy of the GOI encouraged the inflow of fresh foreign capital by allowing 100 % foreign equity in certain projects under the automatic route. NRIs and OCBs (Overseas Corporate Bodies) may invest up to 100 % capital with reparability in high priority industries. MNCs and TNCs were encouraged to establish themselves in Indian markets and were given a level playing field to compete with Indian enterprises.

3. Foreign Exchange Regulation Act (FERA) was liberalized in 1993 and later Foreign Exchange Management Act (FEMA) 1999 was passed to enable foreign currency transactions. India signed many agreements with the WTO affirming it's commitment to liberalize trade such

as TRIPs (Trade Related Intellectual Property Rights), TRIMs (Trade Related Investment Measures) and AOA (Agreement On Agriculture).

Impact of Globalization:

Advantages of Globalization:

- There is a decline in the number of people living below the poverty line in developing countries due to increased investments, trade and rising employment opportunities.
- There is an improvement in various economic indicators of the LDCs (Less Developed Countries) such as employment, life expectancy, literacy rates, per capita consumption etc.
- Free flow of capital and technology enables developing countries to speed up the process of industrialization and lay the path for faster economic progress.
- Products of superior quality are available in the market due to increased competition, efficiency and productivity of the businesses and this leads to increased consumer satisfaction.
- Free flow of finance enable the banking and financial institutions in a country to fulfill financial requirements through internet and electronic transfers easily and help businesses to flourish.
- MNCs bring with them foreign capital, technology, know-how, machines, technical and managerial skills which can be used for the development of the host nation.

Disadvantages of Globalisation:

- Domestic companies are unable to withstand competition from efficient MNCs which have flooded Indian markets since their liberalized entry. It may lead to shut down of operations, pink slips and downsizing. Moreover skilled and efficient labour get absorbed by these MNCs that offer higher pay and incentives leaving unskilled labour for employment in the domestic industries. Thus there may be unemployment and underemployment.
- Payment of dividends, royalties and repatriation has in fact led to a rise in the outflow of foreign capital.
- With increased dependence on foreign technology, development of indigenous technology has taken a backseat and domestic R and D development has suffered.
- Globalization poses certain risks for any country in the form of business cycles, fluctuations in international prices, specialization in few exportables and so on.
- It increases the disparities in the incomes of the rich and poor, developed nations and LDCs. It leads commercial imperialism as the richer nations tend to exploit the resources of the poor nations. Globalization leads to fusion of cultures and inter-mingling of societies to such an extent that there may be a loss of identities and traditional values. It gives rise to mindless aping of western lifestyles and mannerisms however ill-suited they may be.
- It leads to overcrowding of cities and puts pressure on the amenities and facilities available in urban areas.

2.12 Digital Divide

The term digital divide was introduced in the mid-1990s and defined as the gap separating those who have access to new forms of information technology from those who do not. The digital divide remains an important public policy debate that encompasses social, economic and political issues (Srinuan and Bohlin, 2011).

In the early 1990's, Tim Bemes-Lee developed the global hypertext system, the World Wide Web, with an aim to provide a common space where information could be shared without barriers. The expansion of the Web may have surprised even its creator. In less than ten years, the online population has grown to 180 million individuals across all continents, while an estimated 250,000 sites are added to the Web each month. Rapid expansion is not unique to the Web. Computers, a strange word some fifty years ago, are now common household items and integral parts of educational systems in many countries. At the end of 1998, more than 40 percent of the households in the United States owned computers and one fourth had Internet access.

In October 1999, ninety per cent of all Canadian schools were online; four out of ten students had used e-mail during the previous school year; and 30 per cent had designed their own web sites. Scholars, journalists and practitioners reacted to the rapid development of the new information and communication technologies (ICTs) with high expectations and equally great concerns. All recognize the technology's potential to overcome geographical and cultural barriers and bring needed improvement to people's lives all over the world. At the same time, fears have mounted that this potential is not being tapped. Instead of fostering a new equilibrium among countries, the ICT revolution may be widening the gap between the "haves" and the "have-nots," and creating a divide that may prove extremely difficult to close.

A digital divide is an economic and social inequality with regard to access to, use of, or impact of information and communication technologies (ICT). The divide may refer to inequalities between individuals, households, businesses, or geographic areas, usually at different socioeconomic levels or other demographic categories. The divide among different countries or regions of the world is referred to as the global digital divide. Traditionally the nature of the divide has been measured in terms of the existing numbers of subscriptions and digital devices. Given the increasing number of such devices, some have concluded that the digital divide among individuals has increasingly been closing as the result of a natural and almost automatic process.

The global digital divide is a term used to describe ‘great disparities in opportunity to access the Internet and the information and educational/business opportunities tied to this access, between developed and developing countries’. Unlike the traditional notion of the "digital divide" between social classes, the "global digital divide" is essentially a geographical division. The 'global digital divide' is distinguishable from the 'digital divide', a phenomenon wherein the rich get richer and the poor get poorer, at least with respect to technology, as the gap between the technological haves and have-nots widens.

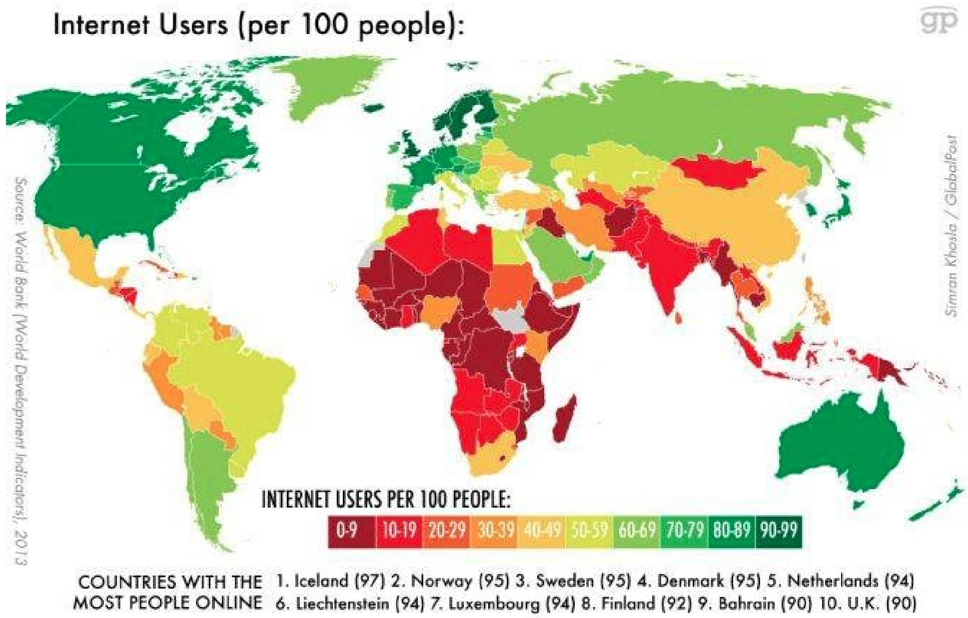


Figure: Digital Divide – Percentage of Internet Users in the World
Source: https://medium.com/@spencerEID100/

Reasons of Digital Divide

The disconnected are not randomly distributed, but have specific demographic, social, economic, racial, ethnic, gender, and political characteristics that amount to a systematic bias of exclusion, often referred to as the "digital have-nots". Similarly, the connected are not randomly distributed, but possess particular demographic, social, economic and political characteristics making up what has become known as the Digital Haves. The separation, chasm, abyss, canyon, gulf, or distance between the Digital Haves and digital have-nots has become known as the Digital Divide.

- Economic disparity
- Educational disparity
- Demographic differences
- Differences in e-learning levels

Overcoming the Divide

Reducing the digital divide is often considered as a major political matter in the digital economy. In fact, that very huge blurred notion encompasses many different questions mixing social, economic, and spatial dimensions.

An individual must be able to connect in order to achieve enhancement of social and cultural capital as well as achieve mass economic gains in productivity. Therefore, access is a necessary (but not sufficient) condition for overcoming the digital divide. Access to ICT meets significant challenges that stem from income restrictions.Social media websites serve as both manifestations of and means by which to combat the digital divide. The former describes phenomena such as the divided users' demographics that make up sites such as Facebook and Myspace or Word Press and Tumblr. Each of these sites host thriving communities that engage with otherwise marginalized populations. An example of this is the large online community devoted to Afrofuturism, a discourse that critiques dominant structures of power by merging themes of science fiction and blackness. Social media brings together minds that may not otherwise meet, allowing for the free exchange of ideas and empowerment of marginalized discourses.

2.13Concept of Ring road, By-pass, Golden Quadrilateral, North-South and East-West Corridor

9.1. Ring road

A ring road (also known as beltline, beltway, circumferential (high)way, loop or orbital) is a road or a series of connected roads encircling a town, city, or country. The most common purpose of a ring road is to assist in reducing traffic volumes in the urban centre, such as by offering an alternate route around the city for drivers who do not need to stop in the city core.The name "ring road" is used for the majority of metropolitan circumferential routes.

Example

Ring Road comprising Rajarhat Expressway (feeds the ring road), Barasat Bypass (north-western orbital) and Eastern Expressway, Eastern Metropolitan Bypass (Eastern orbital, not a motorway), Kolkata, West Bengal

The Eastern Metropolitan Bypass (E.M. Bypass, EM Bypass) is a 29 kilometres (18 mi) major road on the east side of Kolkata. It connects Ultadanga in Bidhannagar (Salt Lake) (northeast Kolkata) to Kamalgazi in RajpurSonarpur (South Kolkata). The road is a major link to Salt Lake and Rajarhat. A lot of high-end construction and development in Kolkata has been centred along the Bypass.

Figure: Outer and Inner Ring Roads of Kolkata Metropolitan Region
Source: http://vaichatt.blogspot.com/

9.2. By-pass

A bypass is a road or highway that avoids or 'bypasses' a built-up area, town, or village, to let through traffic flow without interference from local traffic, to reduce congestion in the built-up area, and to improve road safety. Bypass routes are often controversial, as they require the building of a road carrying heavy traffic where no road previously existed.

Objectives

Highway improvements such as bypass construction typically are motivated by a desire to improve the flow and safety of travel. But, given the importance of travel, transportation improvement projects often can affect the local economy and quality of life.

Example – EM By-pass in Kolkata Metropolitan Area

9.3. Golden Quadrilateral

The Golden Quadrilateral is a highway network connecting many of the major industrial, agricultural and cultural centres of India. It forms a quadrilateral connecting Chennai, Kolkata, Delhi and Mumbai.

It is the largest highway project in India and the fifth longest in the world, at 5,846 km. Started by the NDA government led by Lt. Prime Minister AtalBihari Vajpayee. It is the first phase of the National Highways Development Project (NHDP), and consists of 5,846 km (3,633 mi) four/six lane express highways, built at a cost of 600 billion (US$8.7 billion). The project was launched in 2001 by Lt. AtalBihari Vajpayee, and was completed in 2012. The vast majority of system is not access controlled, although safety features such as guardrails, shoulders, and high-visibility signs are in use.

Figure: Golden Quadrilateral Highway of India

Source: Prepared by Information Technology, Planning Division, NHAI and accessed from

https://www.campustimespune.com/

Table: Important cities connected by Golden Quadrilateral highway

Delhi – Kolkata	Kolkata – Chennai	Chennai – Mumbai	Mumbai – Delhi
• Delhi	• Kolkata	• Chennai	• Mumbai
• Faridabad	• Kharagpur	• Sriperumbudur	• Silvassa
• Mathura	• Balasore	• Kanchipuram	• Vapi
• Agra	• Cuttack	• Walajapet	• Valsad
• Firozabad	• Bhubaneswar	• Ranipet	• Navsari
• Etawah	• Visakhapatnam	• Vellore	• Surat
• Kanpur	• Rajahmundry	• Pallikonda	• Bharuch
• Fatehpur district	• Eluru	• Ambur	• Ankleshwar
• Allahabad	• Vijayawada	• Vaniyambadi	• Vadodara
• Varanasi	• Guntur	• Krishnagiri	• Anand
• Chandauli	• Nellore	• Hosur	• Nadiad
• Mohania	• Chennai	• Bengaluru	• Ahmedabad
• Kudra		• Tumakuru	• Gandhinagar
• Sasaram		• Sira	• Udaipur
• Dehri		• Chitradurga	• Chittaurgarh
• Aurangabad		• Davangere	• Ajmer
• Sherghati		• Ranebennur	• Jaipur
• Dobhi		• Hubballi-Dharwad	• Gurgaon
• Chouparan		• Belagavi	• Delhi
• Barhi		• Kolhapur	
• Ishri		• Sangli-Miraj	
• Bagodar		• Karad	
• Dhanbad		• Satara	
• Asansol		• Pune	
• Durgapur		• Panvel	
• Bardhaman		• Mumbai	
• Kolkata			

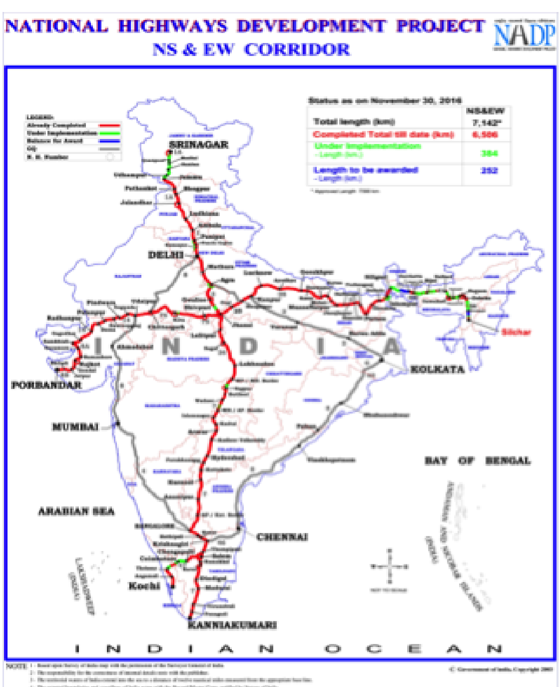


Figure: NS & EW Corridor

9.4. North-South and East-West Corridor

The North–South–East–West Corridor (NS-EW) is the largest on-going highway project in India. The North–South Corridor connects Srinagar with Kanyakumari. The East–West Corridor extends from Porbandar to Silchar.Jhansi is the junction of north-south and east-west corridor.North- South corridor length is 4000 km whereas east west corridor length is 3300 km.As of April 2012, 84.26% of the project had been completed and 15.7% of the project work is currently at progress. It also includes Port connectivity and other projects — 1,157 km (719 mi).

Road transport, being the most accessible mode of transport, is vital to the economic development of the nation. Social integration of the nation increases due to easy, safe and efficient road transportation. Road transport has gained higher share of both passenger and freight traffic compared to other modes of transport due to easy accessibility and reliability. To cater the increasing needs of a growing economy like India and increase in demand for transport services, it is required to expand, develop and improve road networks. This task of road development was decided to be done in a planned manner by the Government of India (Somani et al, 2013).

Various road development plans were started by keeping in mind the future needs of India like National Highway Development Plan (NHDP) was started by BJP in 1999, PradhanMantri Gram SadakYojana (PMGSY) was also started by BJP in 2000, improving road connectivity to ports and airports and many more.

Improving the overall safety, promote new technology, considering social and environmental factors and to reduce the fuel cost by improving quality of roads were the important concerns of the planning committee.

2.14Significance of Trade in Regional and National Economy

Trade involves the transfer of goods or services from one person or entity to another, often in exchange for money. A system or network that allows trade is called a market. An early form of trade, barter, saw the direct exchange of goods and services for other goods and services.

In one modern view, trade exists due to specialization and the division of labour, a predominant form of economic activity in which individuals and groups concentrate on a small aspect of production, but use their output in trades for other products and needs. Trade exists between regions because different regions may have a comparative advantage (perceived or real) in the production of some trade-able commodity—including production of natural resources scarce or limited elsewhere, or because different regions' sizes may encourage mass production. In such circumstances, trade at market prices between locations can benefit both locations.

Importance of trade

a. <u>Make use of abundant raw materials</u>

Some countries are naturally abundant in raw materials – oil (Qatar), metals, fish (Iceland), Congo (diamonds) Butter (New Zealand). Without trade, these countries would not benefit from the natural endowments of raw materials.

A theoretical model for this was developed by Eli Heckscher and Bertil Ohlin. Known as the Heckscher–Ohlin model (H–O model) it states countries will specialise in producing and exports goods which use abundant local factor endowments. Countries will import those goods, where resources are scarce.

b. <u>Comparative advantage</u>

The theory of comparative advantage states that countries should specialise in those goods where they have a relatively lower opportunity cost. Even if one country can produce two goods at a lower absolute cost – doesn't mean they should produce everything. India, with lower labour costs, may have a comparative advantage in labour-intensive production (e.g. call centres, clothing manufacture). Therefore, it would be efficient for India to export these services and goods. While an economy like the UK may have a comparative advantage in education and video game production. Trade allows countries to specialise. More details on how comparative advantage can increase economic welfare.

The theory of comparative advantage has limitations, but it explains at least some aspects of international trade.

c. Greater choice for consumers

New trade theory places less emphasis on comparative advantage and relative input costs. New trade theory states that in the real world, a driving factor behind the trade is giving consumers greater choice of differentiated products. We import BMW cars from Germany, not because they are the cheapest but because of the quality and brand image. Regarding music and film, trade enables the widest choice of music and film to appeal to different tastes. When the Beatles went on tour to the US in the 1960s, it was exporting British music – relative labour costs were unimportant.

Perhaps the best example is with goods like clothing. Some clothing (e.g. value clothes from Primark – price is very important and they are likely to be imported from low-labour cost countries like Bangladesh. However, we also import fashion labels Gucci (Italy) Chanel (France). Here consumers are benefitting from choice, rather than the lowest price. Economists argue that international trade often fits the model of monopolistic competition. In this model, the important aspect is brand differentiation. For many goods, we want to buy goods with strong brands and reputations. e.g. popularity of Coca-Cola, Nike, Addidas, McDonalds etc.

d. Specialisation and economies of scale – greater efficiency

Another aspect of new trade theory is that it doesn't really matter what countries specialise in, the important thing is to pursue specialisation and this enables companies to benefit from economies of scale which outweigh most other factors. Sometimes, countries may specialise in particular industries for no over-riding reason – it may just be a historical accident. But, that specialisation enables improved efficiency. For high value-added products, multinationals often split the production process into a global production system. For example, Apple designs their computers in the US but contract the production to Asian factories. Trade enables a product to have multiple country sources. With car production, the productive process is often even more global with engines, tyres, design and marketing all potentially coming from different countries.

e. Service sector trade

Trade tends to conjure images of physical goods import bananas, export cars. But, increasingly the service sector economy means more trade is of invisibles – services, such as insurance, IT services and banking. Even in making this website, I sometimes outsource IT services to developers in other countries. It may be for jobs as small as $50. Furthermore, I may export a revision guide for £7.49 to countries all around the world. A global economy with modern communications enables many micro trades, which wouldn't have been as possible in a pre-internet age.

f. Global growth and economic development

International trade has been an important factor in prompting economic growth. This growth has led to a reduction in absolute poverty levels – especially in south East Asia which has seen high rates of growth since the 1980s.

India's Trade and Capital Flows in Recent Years

Reforms were dramatic during the first few years after 1991, and reform momentum continued, or was renewed, on a number of fronts up until the past several years. While there has been little backsliding, the momentum for reform seems to have been lost shortly after the coalition government led by Prime Minister Manmohan Singh came into office in 2004.

To be sure, some measures have been undertaken. The GOI has entered into a number of preferential trading arrangements [PTA]), primarily with South and South-East Asian nations. Agreements with Sri Lanka and Singapore date from the 1990s, but others are either in the process of negotiation or in the early stages of implementation. India has signed a South Asian Free Trade Agreement (SAFTA) but there are a number of issues remaining to be negotiated. The country is also committed to a preferential arrangement with the Association of South-East Asian Nations (ASEAN). To a considerable extent, India's trade has been relatively geographically diversified, and the decision to enter into PTAs was defensive, as PTAs were proliferating worldwide. The share both of intraregional trade and trade governed by PTAs remains relatively small.

Although India had been a founding member of the GATT/WTO, India's stance had historically been supportive of the multilateral system but insistent upon special treatment for developing countries and resistant to any liberalization. Since the opening up of the Indian economy, India has focused on agriculture and the difficulties India would have in reducing protection for its agriculture in the Doha Round of trade negotiations. India and Brazil have been two key participants from emerging markets in the Doha Round, but the GOI has resisted the lowering of its bound tariffs to the currently applied rates even in manufacturing, and remains a supporter of the developing countries' positions on issues such as special and differential treatment, trade facilitation, and intellectual property rights.

A second recent trade initiative has been the move toward Special Economic Zones (SEZ). The intent was to enable exporters to avoid much of both the bureaucratic red tape governing transactions and the restrictive labor laws. The intent is to promote the development of large-scale manufacturing of unskilled labor-intensive goods. The legislation permitting SEZs was passed in 2005 and regulations for implementation were promulgated early in 2006. It is thus too soon to evaluate their effect. However, in the winter of 2007, licenses granting SEZ status were suspended for several months as strong political objections were raised on the grounds that farmers were losing their lands and that large and/or rich enterprises were using the legislation to obtain land inappropriately.

By the time of writing, the foreign trade regime has been radically reformed from its pre-1991 highly restrictive stance. Tariffs were, as already seen, still high by standards of other emerging markets, but an official target was to bring them down to ASEAN levels in the near future. Moreover, protection levels were very low compared to those existing in the 1980s especially as quantitative restrictions had been almost entirely phased out. However, while tariffs on manufactures had fallen sharply, tariffs and non-tariff barriers on agricultural commodities remained steep, averaging 40% at the time of the last WTO Trade Policy Review (2007).

The removal of quantitative restrictions and huge tariff reductions lessened the restrictiveness of the regime enormously. There had been a partial offset, at least in the early 2000s, as antidumping measures came to be used. Panagariya (2008) presents evidence that between 1995 and 2005, India was the largest user of antidumping measures, with a cumulative 425 initiations. By contrast, the USA (the second largest user) had initiated 366 cases, the European Community 327, and the fourth largest, Argentina, had introduced 204 cases. There is some evidence that resort to antidumping measures has diminished in recent years. Restrictions on foreign providers of services remained largely unchanged. Estimates by the Australian Productivity Commission put trade restrictiveness indices for India at 59.9 for banking, 68.9 for telecoms, and 60.5 for maritime services, with lower but still high numbers for distribution (32.3), accountancy (44.3), and other services.

These contrast with numbers for industrial countries in banking where the highest restrictiveness index number was 19.4 (Japan), and in distribution where France's number was 32.7, Japan's and Germany's were both close to 25, but other countries had much lower numbers. The response in the foreign trade sector was much greater than even the optimists anticipated. As already seen, export earnings rose dramatically, and India was much more integrated into the world economy, although its share of world exports was still low (at 1%) contrasted with its share in the early 1950s. Although there is some indication that export growth may have slowed in the last half of 2007 and early 2008, merchandise export growth averaged over 20% annually from 2003–2004 to 2005–2006, the last year for which data are available. In addition, service exports grew exceptionally rapidly, more than doubling in each of the three most recent years. Indeed, in 2005/06, India had a positive services balance equal to 3% of GDP.

Foreign investment also rose sharply in response to the relaxation of conditions and improved incentives. Annual FDI inflows rose from $3.13 billion in 2002/03 to $5.6 billion in 2005/06, while inward portfolio investment, which had been less than $1 billion as recently as 2002/03, rose to $12.5 billion in 2005/06.

For both trade in services and inward foreign investment, a major development was the rapid growth of the information technology (IT) sector in India. Starting from a very low level in the early 1990s, it increased in importance, as world-class firms, such as Infosys and Wipro emerged, and outsourcing to India became increasingly important for many firms in industrial countries. Expectations are that robust growth of the sector will continue, although the very exceptionally high rate of growth may decline as labor shortages and other bottlenecks are appearing.

IT service exports grew from about $800 million in 1995/96 to $17.3 billion in 2005/ 06 and rapid growth continues. Total software exports were $23.6 billion in 2005–2006, and constituted 39% of total services exports. In turn, India's share of world exports of services was 2.5%, reflecting the achievements in this area.

Some observers have questioned whether the success of Indian IT services was the result of government intervention and government policies. In fact, however, the opposite seems to be the case: the IT sector started its rapid growth period without significant government encouragement. Indeed, a major factor seems to have been that it is a service industry that was not subject to significant government regulation during its early period. Hence, it was able to become established with much less government intervention than was the case with many other industries. In addition, the IT industry was much less dependent on Indian infrastructure than were most other manufactures. Satellite communications were a major vehicle for transmission of product between foreign buyers and Indian providers. The success of the Indian IT sector exceeded all expectations of what India could achieve, and has resulted in a considerable increase in optimism about India's economic capabilities. This last may be the most important contribution of the IT sector to India's economic policy and growth prospects.

2.15 Export Processing Zones

Export processing zones (EPZs) are areas within developing countries that offer incentives and a barrier-free environment to promote economic growth by attracting foreign investment for export-oriented production.The number of zones internationally, countries hosting EPZs, and firms operating in them, and the business volume they handle, are all growing rapidly, suggesting their importance.

Objectives

The basic objective of setting up EPZs in India is to promote exports and foreign exchange earnings. Though the objectives of EPZs were not clearly spelt out in India until the late 1980s, in actual practice the predominant condition in selecting EPZ units had been the expected value addition component of exports.

Brief history of the EPZ Policy in India

Attempts to promote the EPZ as an export platform on the basis of economic incentives, such as the free provision of infrastructural services and tax holidays, has been a feature of Indian development thinking since the 1960s. The country has had four phases in the evolution of the EPZ policy. Following is an overview of the evolution of the EPZ policy in India through these four phases.

- Initial Phase: 1964-1985

The first zone was set up in Kandla as early as 1965. It was followed by the Santacruz export processing zone which came into operation in 1973. There was however no clarity of objectives that the government wanted to achieve. Kandla and Santacruz EPZs were set up with multiple objectives.There was no single window facility within the zone. Entrepreneurs had to acquire individual clearances from various state government and central government departments. Day-to-day operations were subjected to rigorous controls. Custom procedures for bonding, bank guarantees and movement of goods were rigid. FDI policy was also highly restrictive.

- The Expansionary phase : 1985-1991

In its report, the Tondon Committee strongly recommended to locate 4 to 5 more zones in the country to provide a fillip to the country's export promotion efforts. It argued that the excessive protectionism had imparted a significant bias against exports. At the same time, the high cost of production structure created by heavy protection reduced the competitiveness of Indian exports. It suggested that free trade zones, which insulated the export sector from various controls and regulations, could be a useful instrument of export promotion. Following the report, the government decided to establish four more zones in 1984. These were at Noida (Uttar Pradesh), Falta (West Bengal) Cochin (Kerala) and Chennai (Tamil Nadu). Visakhapatnam EPZ in Andhra Pradesh was established in 1989. The Kandla Free Trade Zone, India's first Export Processing Zone was set up in 1965. Subsequently, six more EPZs were set up at Santa Cruz (Mumbai), Falta (West Bengal), Chennai (Tamil Nadu), Noida (UP), Cochin (Kerala), and Visakhapatnam (Andhra Pradesh).

- The Consolidating Phase : 1991-2000

In 1991, a massive dose of liberalization was administered in the Indian economy. In this phase of new policy initiatives, wide-ranging measures were initiated by the government for revamping and restructuring EPZs.This phase was thus marked by progressive liberalisation of policy provisions and relaxation in the severity of controls and simplification of procedures.

- The emergence phase : 2000 onwards

This period has witnessed a major shift in direction, thrust and approach. The EXIM Policy (1997-2002) has introduced a new scheme from April 1, 2000 for establishment of the Special Economic Zones (SEZs) in different parts of the country. SEZs will be permitted to be set up in the public, private, joint sector or by the State Governments with a minimum size of not less than 1000 hectares. SEZ is an almost self-contained area with high class infrastructure for commercial as well as residential inhabitation. The units operating in these zones are to be deemed as outside the country's customs territory and will have full flexibility of operations.

2.16 Exclusive Economic Zones

An exclusive economic zone (EEZ) is a sea zone prescribed by the United Nations Convention on the Law of the Sea over which a state has special rights regarding the exploration and use of marine resources, including energy production from water and wind. It stretches from the baseline out to 200 nautical miles (nmi) from its coast. The state coastal baselines are less than 400 nmi (740 km) apart. When an overlap occurs, it is up to the states to delineate the actual maritime boundary.

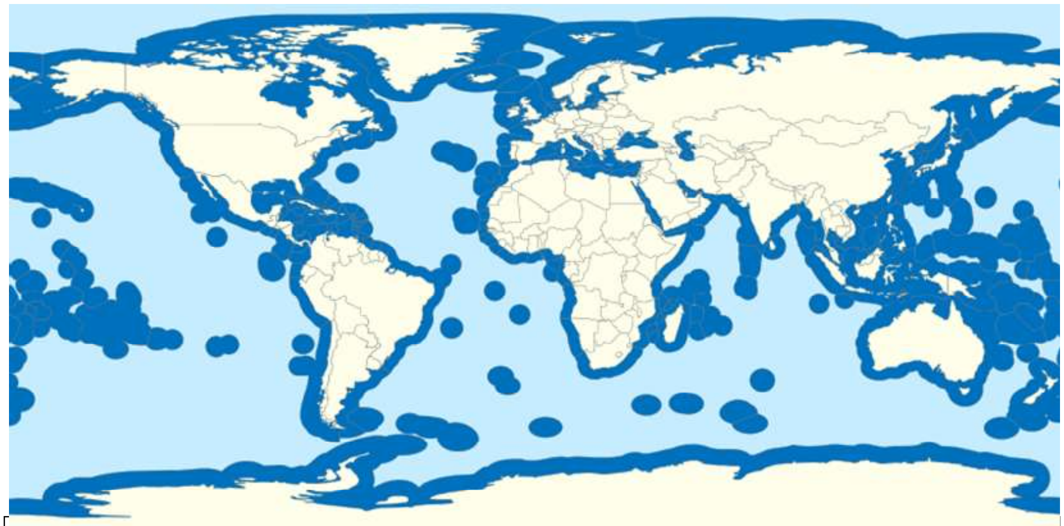


Figure: The World's Exclusive Economic Zones

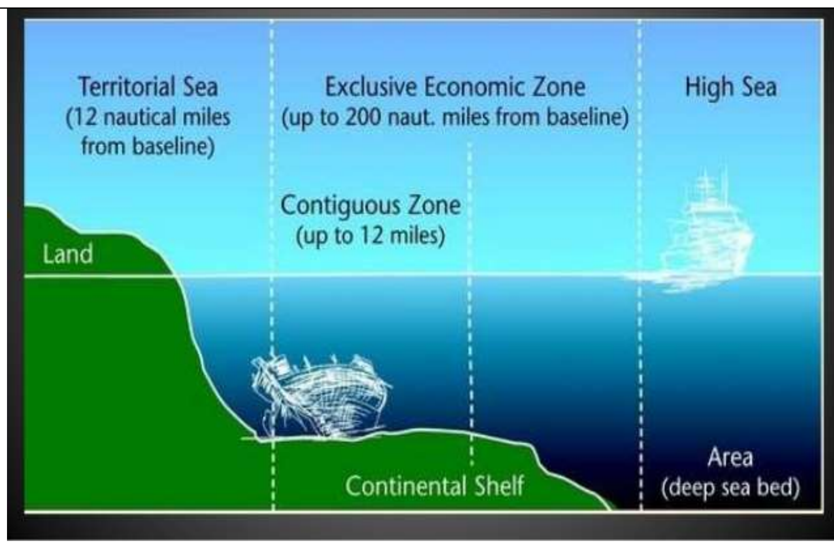


Figure: Demarcation of Exclusive Economic Zone

History

As Patuzi (2015: 149-155) has analysed, the principle of freedom of the seas is recognized in the seventeenth century as one of the oldest principles of international law. This concept of freedom was ideally suited to the requirements of commerce and economic progress and was the sea-going equivalent of the liberal principles of free trade and free enterprise. In December, 1982, the Montego Bay Convention (Jamaica) on the Law of the Sea was signed. The Republic of Albania has approved the Convention of UN on the Law of the Seas with the Albanian Law No. 9055, on 24th April, 2003. The 1982 UN Convention on the Law of the Sea (UNCLOS) was considered as one ofthe most successful and progressive codifications made by the United Nations since the end of the World War II. As a comprehensive legal framework for the law of the sea, the UNCLOS Convention has elucidated the rights and obligations of all States, including: coastal, land-locked and geographical disadvantaged States and other international actors in various functional maritime areas.

The problems addressed in the 1982 Convention can be divided in two major groups:

a) Marine spaces that are subject to national jurisdiction (internal waters, territorial sea, contiguous zone, exclusive economic zone, continental shelf);

b) Marine spaces that are not subject to national jurisdiction. The newest concept of the Montego Bay Convention was the Exclusive Economic Zone.

The EEZ, (Article 55) is an area beyond and adjacent to the territorial sea, subject to the specific legal regime established in this part, under which the rights and jurisdiction of the coastal State and the rights and freedoms of other States are governed by the relevant provisions of this Convention. The main characteristic of the EEZ regime is the fact that the coastal state does not exercise sovereignty in its entire territorial, but only some rights such as exploration, exploitation, conservation and resource management. The jurisdiction of the Coastal state is about the rise and use of the artificial installations, research into the sea, as well as protection of the marine environment. Between freedom of navigation and the immediate interest of the coastal state, priority is given to the coastal State.

Rights and duties of the Coastal State in the EEZ

According to Article 56, the coastal State has sovereign rights in the exclusive economic zone for the purpose of exploring and exploiting, conserving and managing the natural resources, whether living or non-living, of the waters superjacent to the seabed and of the seabed and its subsoil, and with regard to other activities for the economic exploitation and exploration of the zone, such as the production of energy from the water, currents and winds. The jurisdiction as provided for in the relevant provisions of this Convention with regard to the establishment uses the artificial islands, installations and structures, marine scientific research and the protection and preservation of the marine environment.

The coastal State shall have exclusive jurisdiction over such artificial islands, installations and structures, including jurisdiction with regard to customs, fiscal, health, safety and immigration laws and regulations. Due notice must be given to the construction of such artificial islands, installations or structures, and permanent means giving warning of their presence. Any installations or structures which are abandoned or disused shall be removed to ensure safety of navigation, taking into account any generally accepted international standards established in this regard by the competent international organization. Such removal shall also have due regard to fishing, protection of the marine environment and the rights and duties of other States. Appropriate publicity shall be given to the depth, position and dimensions of any installations or structures not entirely removed.

Right of land-locked States

Land-locked States expressed in the Article 69, shall have the right to participate, on an equitable basis, in the exploitation of an appropriate part of the surplus of the living resources of the exclusive economic zones of coastal States of the same sub region or region, taking into account the relevant economic and geographical circumstances of all the States concerned and in conformity with the provisions of this article. The terms and modalities of such participation shall be established by the States concerned through bilateral, sub regional or regional agreements taking into account, the need to avoid effects detrimental to fishing communities or fishing industries of the coastal State; the extent to which the land-locked State, in accordance with the provisions of this article, is participating or is entitled to participate under existing bilateral, sub- regional or regional agreements in the exploitation of living resources of the exclusive economic zones of other coastal States.

Also those agreements must take into account the extent to which other land-locked States and geographically disadvantaged States are participating in the exploitation of the living resources of the exclusive economic zone of the coastal State and the consequent need to avoid a particular burden for any single coastal State or a part of it. The nutritional needs of the populations of the respective States shall also be taken into account.

When the harvesting capacity of a coastal State approaches a point which would enable it to harvest the entire allowable catch of the living resources in its exclusive economic zone, the coastal State and other States concerned shall cooperate in the establishment of equitable arrangements on a bilateral, sub regional or regional basis to allow for participation of developing land-locked States of the same sub region or region in the exploitation of the living resources of the exclusive economic zones of coastal States of the sub region or region, as may be appropriate in the circumstances and on terms satisfactory to all parties.

To implement this provision, developed land-locked States shall be entitled to participate in the exploitation of living resources only in the exclusive economic zones of developed coastal States of the same sub region or region having regard to the extent to which the coastal State, in giving access to other States to the living resources of its exclusive economic zone, has taken into account the need to minimize detrimental effects on fishing communities and economic dislocation in States whose nationals have habitually fished in the zone. The above provisions are without prejudice to arrangements agreed upon in sub regions or regions where the coastal States may grant to land-locked States of the same sub region or region equal or preferential rights for the exploitation of the living resources in the exclusive economic zones.

Right of geographically disadvantaged States

"Geographically disadvantaged States" mean coastal States, including States bordering enclosed or semi-enclosed seas, whose geographical situation makes them dependent upon the exploitation of the living resources of the exclusive economic zones of other States in the sub region or region for adequate supplies of fish for the nutritional purposes of their populations or parts thereof, and coastal States which can claim no exclusive economic zones of their own.

Geographically disadvantaged States (Article 70) shall have the right to participate, on an equitable basis, in the exploitation of an appropriate part of the surplus of the living resources of the exclusive economic zones of coastal States of the same sub region or region, taking into account the relevant economic and geographical circumstances of all the States concerned and in conformity with the provisions of this article and articles of the Convention. The terms and modalities of such participation shall be established by the States concerned through bilateral, sub regional or regional agreements taking into account:

(a) the need to avoid effects detrimental to fishing communities or fishing industries of the coastal State;

(b) the extent to which the geographically disadvantaged State, in accordance with the provisions of this article, is participating or is entitled to participate under existing bilateral, sub regional or regional agreements in the exploitation of living resources of the exclusive economic zones of other coastal States;

(c) the extent to which other geographically disadvantaged States and land-locked States are participating in the exploitation of the living resources of the exclusive economic zone of the coastal State and the consequent need to avoid a particular burden for any single coastal State or a part of it;

(d) the nutritional needs of the populations of the respective States.

Significance of EEZ

- **Fishery Development**

Fishery development received much more emphasis during the 1980s. First, the worldwide acceptance of the concept of Exclusive Economic Zones in the sea out to 200 miles from the coast (called the Fishery Conservation Zone in the United States) led to the perception by coastal countries that they had wealth in their EEZ ripe for the taking. Second, the higher price of fish resulting from the levelling off of marine production led to increased interest in fish farming and increased production. Third, the rapidly increasing interest in recreational fishing is leading to allocation of resources to, and government development of, recreational fisheries, sometimes at the expense of commercial fisheries.

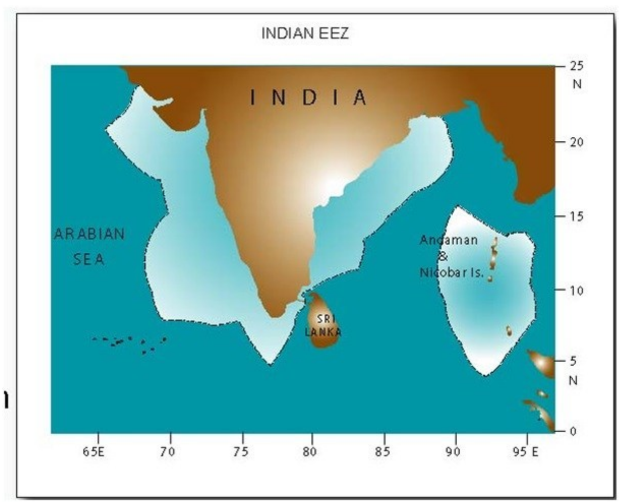
- **Environmental Protection and Ecology**

Under international law, states have the sovereign right to exploit, manage, and conserve the natural resources and natural systems within their jurisdiction, including resources located in their territorial sea and exclusive economic zone, and sinks such as the atmosphere. States also have a broad right to engage in fishing on the high seas. However, the expansion of the world economy has placed increasing pressure on natural systems that overlap or transcend political boundaries. This has gradually led to the development of a large number of international agreements, systems, and processes to address transnational environmental issues.

- **Mineral Extraction**

Ocean floor minerals are not owned by a mining company. If the minerals are within the exclusive economic zone (EEZ) of a country, they are the property of that country's government.

India's Exclusive Economic Zone



The size of the Indian EEZ is estimated at 2.02-2.2 million sq. km, covering both the western and eastern coasts, as well as the island territories of Lakshadweep in the Arabian Sea and the Andaman and Nicobar Islands in the Bay of Bengal. The variance of 180,000 sq. km is due to the absence of published baselines of the country (from which maritime zones are calculated), as well as a series of minor hydro-graphic differences. Moreover, the size of the EEZ is expected to increase even further by the year 2004, in view of the legal provision of extending the continental shelf to 350 nm, if preliminary exploration of the extended area is completed by then. This could provide India an additional EEZ of approximately 1.5 million sq. km, if its continental shelf extends well beyond 200 nm. or 100 nm. beyond the 2,500 metre isobath.

In the EEZ, India possesses the following rights and jurisdictions:

- sovereign rights for the purpose of exploring and exploiting, conserving and managing the natural resources of the seabed, subsoil, and the waters;
- rights and jurisdiction with regard to the establishment of artificial islands, installations and structures; exclusive jurisdiction with regard to other activities for the economic exploitation and exploration of the zone, such as the production of energy from the water, currents and winds; and jurisdiction with regard to the preservation of the marine environment, including pollution control.

- Other states could utilize the resources of the EEZ only with the prior permission of the Indian government.
- In terms of warfare, the EEZ is to be considered similar to the high seas, but for the additional obligation to have "due regard to the rights and duties of the coastal state".

2.17 Forward Trading

A forward contract is a commitment to purchase at a future date a given amount of a commodity or an asset at a price agreed on today. A forward contract is a customized contract between two parties to buy or sell an asset at a specified price on a future date. A forward contract can be used for hedging or speculation, although its non-standardized nature makes it particularly apt for hedging.

Consider the following example of a forward contract.

Assume that an agricultural producer has two million bushels of corn to sell six months from now and is concerned about a potential decline in the price of corn. It thus enters into a forward contract with its financial institution to sell two million bushels of corn at a price of $4.30 per bushel in six months, with settlement on a cash basis.

In six months, the spot price of corn has three possibilities:

I. It is exactly $4.30 per bushel. In this case, no monies are owed by the producer or financial institution to each other and the contract is closed.

II. It is higher than the contract price, say $5 per bushel. The producer owes the institution $1.4 million, or the difference between the current spot price and the contracted rate of $4.30.

III. It is lower than the contract price, say $3.50 per bushel. The financial institution will pay the producer $1.6 million, or the difference between the contracted rate of $4.30 and the current spot price.

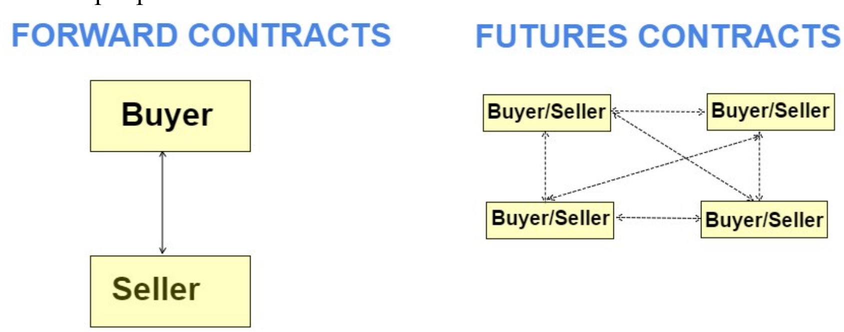


Figure: Forward trading and future trading

Risks with Forward Contracts

The market for forward contracts is huge since many of the world's biggest corporations use it to hedge currency and interest rate risks. However, since the details of forward contracts are restricted to the buyer and seller – and are not known to the general public – the size of this market is difficult to estimate.

The large size and unregulated nature of the forward contracts market mean that it may be susceptible to a cascading series of defaults in the worst-case scenario. While banks and financial corporations mitigate this risk by being very careful in their choice of counterparty, the possibility of large-scale default does exist.

Another risk that arises from the non-standard nature of forward contracts is that they are only settled on the settlement date and are not marked-to-market like futures. What if the forward rate specified in the contract diverges widely from the spot rate at the time of settlement?

In this case, the financial institution that originated the forward contract is exposed to a greater degree of risk in the event of default or non-settlement by the client than if the contract were marked-to-market regularly.

2.18 E-commerce

E-Commerce or Electronics Commerce is a methodology of modern business, which addressesthe need of business organizations, vendors and customers to reduce cost and improve thequality of goods and services while increasing the speed of delivery. E-commerce refers tothe paperless exchange of business information using the following ways:

- Electronic Data Exchange (EDI)
- Electronic Mail (e-mail)
- Electronic Bulletin Boards
- Electronic Fund Transfer (EFT)
- Other Network-based technologies

Even today, some considerable time after the so called ‘dot com/Internet revolution’, electronic commerce (e-commerce) remains a relatively new, emerging and constantly changing area of business management and information technology. There has been and continues to be much publicity and discussion about e-commerce. Library catalogues and shelves are filled with books and articles on the subject. However, there remains a sense of confusion, suspicion and misunderstanding surrounding the area, which has been exacerbated by the different contexts in which electronic commerce is used, coupled with the myriad related buzzwords and acronyms. This book aims to consolidate the major themes that have arisen from the new area of electronic commerce and to provide an understanding of its application and importance to management.

Features

E-Commerce provides the following features:

- Non-Cash Payment: E-Commerce enables the use of credit cards, debit cards, smartcards, electronic fund transfer via bank's website, and other modes of electronicspayment.

- 24x7 Service availability: E-commerce automates the business of enterprises andthe way they provide services to their customers. It is available anytime, anywhere.

- Advertising/Marketing: E-commerce increases the reach of advertising of productsand services of businesses. It helps in better marketing management ofproducts/services.

- Improved Sales: Using e-commerce, orders for the products can be generatedanytime, anywhere without any human intervention. It gives a big boost to existingsales volumes.

- Support: E-commerce provides various ways to provide pre-sales and post-salesassistance to provide better services to customers.

- Inventory Management: E-commerce automates inventory management. Reportsget generated instantly when required. Product inventory management becomes veryefficient and easy to maintain.

- Communication improvement: E-commerce provides ways for faster, efficient,reliable communication with customers and partners.

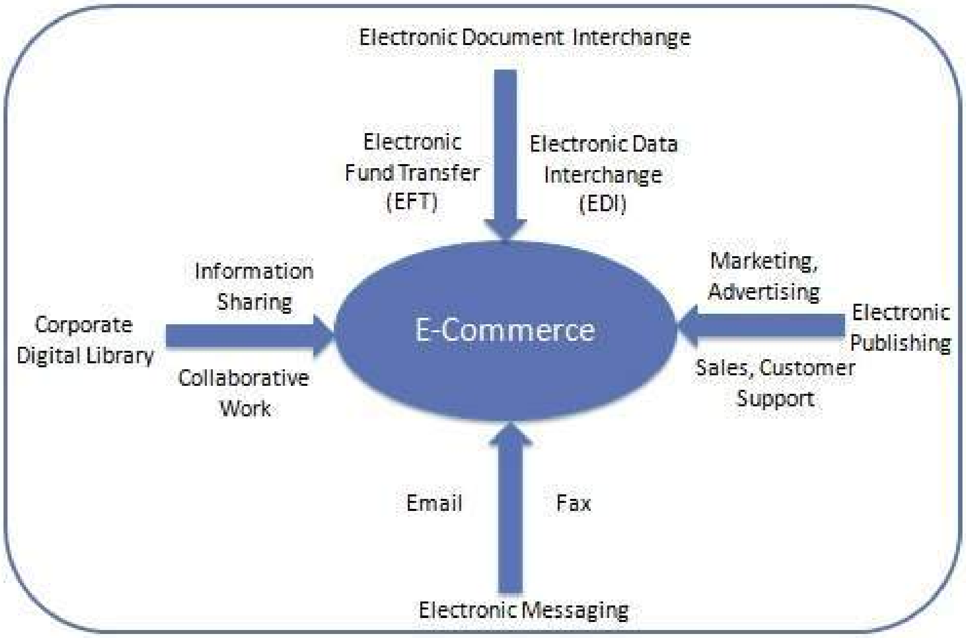


Figure: E-commerce

The advantages of e-commerce can be broadly classified into three major categories:

- **<u>Advantages to Organizations</u>**
  Using e-commerce, organizations can expand their market to national and international markets with minimum capital investment.
   E-commerce helps organizations to reduce the cost to create process, distribute, retrieve and manage the paper based information by digitizing the information.
   E-commerce improves the brand image of the company.
   E-commerce helps organizations to provide better customer service.
   E-commerce helps to simplify the business processes and makes them faster and efficient.
   E-commerce reduces the paper work.
   E-commerce increases the productivity of organizations. It supports "pull" type supply management. In "pull" type supply management, a business process starts when a request comes from a customer and it uses just-in-time manufacturing way.
- **<u>Advantages to Consumers</u>**
  It provides 24x7 support. Customers can enquire about a product or service and place orders anytime, anywhere from any location.
  E-commerce application provides users with more options and quicker delivery of products.
   E-commerce application provides users with more options to compare and select the cheaper and better options.
  A customer can put review comments about a product and can see what others are buying, or see the review comments of other customers before making a final purchase.
   E-commerce provides options of virtual auctions.
  It provides readily available information. A customer can see the relevant detailed information within seconds, rather than waiting for days or weeks.
  E-Commerce increases the competition among organizations and as a result, organizations provide substantial discounts to customers.
- **<u>Advantages to Society</u>**

Customers need not travel to shop a product, thus less traffic on road and low air pollution.
 E-commerce helps in reducing the cost of products, so less affluent people can also afford the products.
 E-commerce has enabled rural areas to access services and products, which are otherwise not available to them.
E-commerce helps the government to deliver public services such as healthcare, education, social services at a reduced cost and in an improved manner.

E-Commerce ─ Disadvantages

The disadvantages of e-commerce can be broadly classified into two major categories:

Technical disadvantages
Non-technical disadvantages

Technical Disadvantages

Therœe can be lack of system security, reliability or standards owing to poor implementation of e-commerce.
 The software development industry is still evolving and keeps changing rapidly.
 In many countries, network bandwidth might cause an issue.
 Special types of web servers or other software might be required by the vendor, setting the e-commerce environment apart from network servers.
 Sometimes, it becomes difficult to integrate an e-commerce software or website with existing applications or databases.
 There could be software/hardware compatibility issues, as some e-commerce software may be incompatible with some operating system or any other component.

Non-Technical Disadvantages

 Initial cost: The cost of creating/building an e-commerce application in-house may be very high. There could be delays in launching an e-Commerce application due to mistakes, and lack of experience.
 User resistance: Users may not trust the site being an unknown faceless seller. Such mistrust makes it difficult to convince traditional users to switch from physical stores to online/virtual stores.
 Security/ Privacy: It is difficult to ensure the security or privacy on online transactions.
 Lack of touch or feel of products during online shopping is a drawback.
 E-commerce applications are still evolving and changing rapidly.
 Internet access is still not cheaper and is inconvenient to use for many potential customers, for example, those living in remote villages.

What is the impact of Electronic Commerce?

E-commerce and e-business are not solely the Internet, websites or dot comcompanies. It is about a new business concept that incorporates all previous business management and economic concepts. As such, e-businessand e-commerce impact on many areas of business and disciplines ofbusiness management studies.
For example:

● **Marketing –**issues of on-line advertising, marketing strategies and consumer behaviour and cultures. One of the areas in which it impacts particularly is direct marketing. In the past this was mainly door-to-door, home parties (like the Tupperware parties) and mail order using catalogues or leaflets. This moved to telemarketing and TV selling with the advances in telephone and television technology and finally developed into e-marketing spawning 'eCRM' (customer relationship management) data mining and the like by creating new channels for direct sales and promotion.

● **Computer sciences** – development of different network and computing technologies and languages to support e-commerce and e-business, for example linking front and back office legacy systems with the 'web-based' technology.

● **Finance and accounting** – on-line banking; issues of transaction costs; accounting and auditing implications where 'intangible' assets and human capital must be tangibly valued in an increasingly knowledge based economy.

● **Economics** – the impact of e-commerce on local and global economies; understands the concepts of a digital and knowledge-based economy and how this fits into economic theory.

● **Production and operations management** – the impact of on-line processing has led to reduced cycle times. It takes seconds to deliver digitized products and services electronically; similarly the time for processing orders can be reduced by more than 90 per cent from days to minutes. Production systems are integrated with finance marketing and other functional systems as well as with business partners and customers (see Intel mini-case).

Production and operations management (manufacturing) – moving from mass production to demand-driven, mass customisation customer pull rather than the manufacturer push of the past. Web-based Enterprise Resource Planning systems (ERP) can also be used to forward orders directly to designers and/or production floor within seconds, thuscutting production cycle times by up to 50 per cent, especially when manufacturing plants, engineers and designers are located in different countries. In sub-assembler companies, where a product is assembled from a number of different components sourced from a number of manufacturers, communication, collaboration and coordination are critical – so electronic bidding can yield cheaper components and having flexible and adaptable procurement systems allows fast changes at a minimum cost so inventories can be minimised and money saved.

● **Management information systems** – analysis, design and implementation of e-business systems within an organisation; issues of integration of front-end and back-end systems.

● **Human resource management** – issues of on-line recruiting, home working and 'intrapreneurs' working on a project by project basis replacing permanent employees.

● **Business law and ethics** – the different legal and ethical issues that have arisen as a result of a global 'virtual' market.Issues such as copyright laws, privacy of customer information, legality of electronic contracts, etc.

Customer Relationship Management (CRM)

Customer relationship management (CRM) focuses on the provision and maintenance of quality service for customers. Effective CRM involves communicating with customers and delivering products, services, information and solutions in response to their problems, wants and needs. Customer satisfaction is key to business success, because it is far less expensive to keep current customers than to acquire new ones. Online businesses should give particular attention to CRM, because transactions are often conducted through a series of third parties, and thus the establishment of personal relationships with customers requires innovative strategies.

Aspects of CRM are call handling (the maintenance of outbound and inbound calls from customers and service representatives), sales tracking (the tracking and recording of all sales made) and transaction support (support for technology and personnel involved in business transactions). Unique functions of eCRM, the application of CRM to an e-business strategy, include the personalization and customization of customers' experiences and interactions with a Web site, call centre or any other forum for customer contact with the e-business. The term iCRM (Internet customer relationship management) can be used interchangeably with eCRM in reference to e-business customer relationship management. Business analysts should review all CRM plan details and data, such as reductions in costs or an influx of customer complaints, to refine the CRM system.

Online Payments

Secure electronic funds transfer (EFT) is crucial to e-commerce. Credit-card payments, digital cash and e-wallets, smart cards, micropayments and electronic bill presentment and payment are methods for conducting online transactions. Many companies offer products, software and services that enable monetary transactions on the Web.

Credit-Card Payment

Although credit cards are a popular form of online payment, many people resist online credit-card transactions because of security concerns. Customers fear credit-card fraud by merchants and third parties. However, most credit cards, such as Visa®, Mastercard® and American Express®, have features that enable secure online and offline payments.

To accept credit-card payments, a merchant must have a merchant account with a bank. Traditional merchant accounts accept only **point-of-sale (POS)** transactions, that is, transactions that occur when customers present credit cards at stores. However, the growth of e-commerce has resulted in the establishment of specialized Internet merchant accounts that handle online credit-card transactions. These consist of **card-not-present (CNP)** transactions. For example, when users make credit-card purchases through the Internet, they can provide the card numbers and expiration dates, but the merchant does not see the actual cards involved in the transactions. A merchant account can be established through either a bank or a third-party service.

Digital Cash and e-Wallets

Digital cash is one example of digital currency. It is stored electronically and can be used to make online electronic payments. Digital-cash accounts are similar to traditional bank accounts; consumers deposit money into digital-cash accounts for use in digital transactions. Often, digital cash is used in conjunction with other payment technologies, such as digital wallets. In addition to providing a payment alternative for customers with security concerns regarding online credit-card transactions, digital cash allows people who do not have credit cards to shop online (see the PayPal feature).

Micropayments

Merchants are required to pay a fee for each credit-card transaction they process, which becomes costly when customers purchase inexpensive items. Sometimes, the cost of an item is actually lower than the standard transaction fee, causing the merchant to incur losses. Micropayments (payments that generally do not exceed $5) enable ways for nominally priced products and services (e.g., music, pictures, texts or videos) to be sold profitably over the Web. For instance, a phone bill is essentially an aggregation of micropayments that are charged periodically at set intervals to justify the transaction fee. To offer micropayment processing, some companies have formed strategic partnerships with telephone carriers and utility companies. In 2002, content purchased through micropayments grew faster than any other category of fee-based content, as sales grew ten times over the previous year to $3 million, and spending for content in the third quarter of 2002 alone was up 132% over the preceding year to $56 million.

Smart Cards

A smart card generally looks like a credit card and can serve many different functions, from authentication to data storage. The most popular smart cards are memory cards and microprocessor cards. Memory cards are similar to floppy disks. Microprocessor cards are similar to small computers, with operating systems, security and storage.Smart cards also have different interfaces with which they interact with reading devices. One type of interface is a contact interface, in which a smart card is inserted into a reading device and physical contact between the device and the card is necessary (For example American Express®'s Blue). The alternative to this method is a contactless interface, in which data is transferred to a reader via an embedded wireless device in the card, without the card and the device having to make physical contact.

Plate: Use of Smart Card in Android Mobile

Contactless cards are convenient for transportation services, such as automatic toll payments. A contactless smart card, when placed in a device in a car, will charge the user's account as he or she drives through toll booths (such as FAST LANE®, E-ZPassSM used in Massachusetts and New York, respectively and FasTRAK™ used in California).25 Smart cards store credit-card numbers, personal contact information, and so on. Each smart card is used in combination with a personal identification number (PIN). This application provides two levels of security by requiring the user to both possess a smart card and know the corresponding PIN to access the information stored on the card. As an added measure of security, some microprocessor cards will delete or corrupt stored data if malicious attempts at tampering with the card occur. In addition, smart cards can require users to enter passwords, thus offering a higher level of security than credit cards. Information maintained on smart cards can be designated as "read only" or as "no access." The cards can also be enhanced with additional security features, such as encryption and photo identification.

Digital signatures

The electronic equivalent of written signatures, are used in public key cryptography to solve authentication and integrity problems. A digital signature authenticates the sender's identity, and, like a written signature, it is difficult to forge. To create a digital signature, a sender first runs a plaintext message through a hash function, which is a mathematical calculation that gives the message ahash value. For example, you could take the plaintext message "Buy 100 shares of company X," run it through a hash function and get a hash value of 42. The hash function could be as simple as adding up all the 1s in a message, although it is usually more complex. The hash value is also known as a message digest. The chance that two different messages will have the same message digest is statistically insignificant. Collision occurs when multiple messages have the same hash value. However, it is computationally infeasible to compute a message from its hash value or to find two messages with the same hash value for hash algorithms commonly used today.

Either the public key or the private key can be used to encrypt or decrypt a message. For example, if a customer uses a merchant's public key to encrypt a message, only the merchant can decrypt the message, using the merchant's private key. Thus, the merchant's identity can be authenticated, since only the merchant knows the private key. However, the merchant has no way of validating the customer's identity, since the encryption key the customer used is publicly available.

2.19 Freight Equalization Policy on Indian Trade

A common goal of industrial policy is to promote balanced economic development, both across sectors and across regions. India's Second Five Year Plan in 1956 made this goal explicit: "Only by securing a balanced and co-ordinated development of the industrial and the agricultural economy in each region, can the entire country attain higher standards of living". Balanced development might prove difficult to attain, however, if some regions are inherently more productive than others. Even if balanced development is feasible, it might prove undesirable, if concentrating economic activity leads to gains from specialization or agglomeration externalities. Indeed, policies such as special economic zones and subsidies for industrial centers aim at precisely the opposite of India's goal: concentration rather than balance.

India enacted the Freight Equalization Scheme (FES) in 1956, with the goal of achieving balanced industrial development. As Figure 1a shows, manufacturing output at the time was heavily concentrated into just a few areas, with West Bengal and the surrounding region being one of the most important centers of production. In 1950, West Bengal and Bihar accounted for 92 percent of all 4 iron and steel production in India and 48 percent of all manufacturing output in engineering-related industries. These areas enjoyed a natural advantage in manufacturing, due to their proximity to raw materials, particularly iron ore (Firth and Liu, 2018).

They were also rich in coal and other important mineral resources. FES served to neutralize this geographic advantage. Starting in 1956, the government fixed uniform prices for the transport of iron and steel. This acted as a subsidy to long-distance shipping, with a user located all the way across the country now being able to obtain iron and steel at the same cost as a user located nearby the sources of materials in West Bengal and Bihar. Along with the equalization of iron and steel to achieve geographic balance in manufacturing, the government also equalized the shipping costs of cement and fertilizers, in order to promote balance in construction and agriculture. Below we find that the cement and fertilizer equalization had little effect on manufacturing activity, and moreover that the data show manufacturing firms making little use of these materials.

So we maintain focus on the equalization of iron and steel.2 The administration of FES for iron and steel fell under the authority of the Ministry of Steel, as detailed by Raza and Aggarwal (1986), Singh (1989), and Mohanty (2015). To finance FES, the government calculated an ex-factory "retention price", which depended on the expected average distance of shipments for the particular type of iron or steel. A self-financing Equalization Fund collected the difference between this price and the actual shipping cost of short shipments, and paid out the associated credits for longer shipments. The fund was administered initially by the government Tariff Commission, then starting in 1964 by a Joint Plant Committee (JPC) established by the Ministry of Steel specifically for price regulation.

The scope of FES was limited to the output of India's integrated steel plants (ISPs). These ISPs in the context of a stylized depiction of the supply chain for iron and steel products in India. The most basic natural resource needed for these products is iron ore, which as noted above, is located primarily in a handful of states in eastern India. The next step in the supply chain is transforming iron ore into basic iron and steel "materials" such as pig iron, structural steel, coils, sheets, and plates. This transformation generally happens at the ISPs, which are controlled by the Ministry of Steel and operate at tremendous scale, giving them a virtual monopoly on the production of the basic materials. There are only seven ISPs in India, and the newest of them was constructed in 1971. So the empirical analysis will regard the location of the ISPs as exogenously fixed, and study where other, more flexible factories choose to produce given these locations.

The users of the materials produced by ISPs can be grouped into two categories. First, makers of processed iron and steel products directly use the ISP output, in order to produce somewhat more specialized products, such as more refined forms of iron and steel, or pieces of iron and steel shaped in particular ways – say into a car axle or chassis. Finally, there is a set of downstream industries using these specialized products, and for example, assembling a car or another final good. Figure 3 indicates how we use industry codes to group industries according to their downstreamness in a way that is consistent across years, including years in the 1950s and 1960s when the industry codes observed in data are coarse.

For recent years with more detailed industry classifications, measures such as Leontief input shares provide a more sophisticated way to characterize the linkages across industries. As of 1956, India had two ISPs in operation, one being the current Tata Steel plant at Jamshedpur, Bihar, and the other being the IISCO plant, operated by the state-owned Steel Authority of India (SAIL) at Burnpur in West Bengal. Figure 4 plots their locations. The concentration of iron and steel material production at these points makes it clear how FES would have served as a subsidy for downstream iron and steel users to move production away from West Bengal and Bihar. After the implementation of FES, several new steel plants opened and fell under the FES scheme: SAIL Bhilai Steel Plant in Madhya Pradesh (now Chhattisgarh), SAIL Rourkela Steel Plant in Orissa, SAIL Bokaro Steel Plant in Bihar (now Jharkhand), and SAIL Durgapur Steel Plant in West Bengal, and Vizag Steel Plant in Andhra Pradesh.

Under FES, the location of these new plants affects average shipping distances and therefore affects the all-India prices of iron and steel products. Since FES keeps prices uniform across regions, however, the new plants do not provide any particular advantage to nearby iron and steel users, and therefore should not affect these users' location choices. The products covered by FES included most iron and steel materials produced by the ISPs for domestic use. Excluded were tin plates, pipes, electrical steel, and alloy steel, though these products amount to a small fraction of the plants' output. The more important products subject to FES included basic materials such as pig iron and steel sheets. These materials serve as inputs to other firms manufacturing more processed or complex products. But the restriction of FES to the ISPs means that the output of these downstream users, even if it contains iron and steel, is still subject to normal, distance-based shipping charges.

FES remained in effect until 1991. The repeal was sudden, with the National Development Council meeting to evaluate FES in December 1991 and announcing its removal with effect from January 1992. In place of FES, the government implemented a "freight ceiling" policy, charging freight based on distance for shorter shipments, but capping the price of longer shipments. The ceiling was, however, set very high: 1125km for pig iron, 1375km for flats, 1400km for bars, and 1500km for semi-steel. In practice, the ceiling did not bind for most users, with only the farthest reaches of northern and southern India lying more than this distance from the nearest steel plant. In 2001, the ceiling was also lifted, marking the complete abolition of FES.

Ex-post appraisals of FES have involved vigorous debate over its effects on the geography of production. A 1977 inter-ministerial report calculated the size of the FES subsidies, concluding that they were inconsequential, amounting to a relatively small fraction of firms' final output prices (Government of India Planning Commission, 1977). Other commentators argue, however, that FES was a driving force for industrial production to move away from eastern India. Over the past halfcentury, the western (and southern) states benefiting from FES have enjoyed India's highest rates of economic growth, while the resource-rich eastern states find themselves among India's poorest states today. Figure 2 shows that the divergence has been especially stark in manufacturing, lending plausibility to the idea that FES and the reversal of the manufacturing advantage contributed to these states' overall reversal of fortune.

Freight Equalization Policy and Sluggish Growth of Indigenous Industries in Eastern India

The Freight Equalization Policy was adopted in 1948 by the Government of India to facilitate the equal growth of industries all over the country. This meant a factory could be set up anywhere in India and the price rate of transportation of some industrial raw materials like coal, iron ore, steel and cement would be subsidized by the Central Government. These materials were available at the same price in all parts of the country. This robbed West Bengal, along with the other States in the Eastern region of India like undivided Bihar and Odisha, of its locational advantage of being endowed with mineral resources like coal and iron ore. Other industrial raw materials like cotton were not coming under the purview of this policy.

This policy induced the growth of heavy and medium scale industries outside the mineral rich regions of the country like Maharashtra, Gujarat and Delhi. After implementation of this policy new factories were set up in those States. According to Ranajit Roy (1971) the railways' own internal rates were 30 for a tonne of steel from Jamshedpur to Haora and 120 for a tonne from Jamshedpur to Mumbai; the total sum was 150 for these two tonnes of steel. Under this policy, the Kolkata user was made to pay 75 and the Mumbai user also 75. The net result was not only that the industrialists of Kolkata to lose his locational advantage of 90 a tonne over his Mumbai competitor but he was also being compelled to pay 45 more for a tonne. On the contrary Mumbai was enjoying the advantage of subsidy at the rate of 45 per tonne. West Bengal and undivided Bihar being main coal producing States of the then India, lost their locational advantages.

The freight (1971 rate) for a tonne of coal from Ranigaunj to Ludhiana is 49.20. But when the same wagons of cotton moved from Fazilka to Haora the freight was 165.70 per tonne; for oilseeds over the same distance the freight was 90.90 per tonne. West Bengal experienced flight of capital and faced an industrial stagnation during the decade of the 1980s as a feedback of this one–sided policy. When the industrialists of the western and northern States of India got the advantages they were interested to invest in their own States. Consequently investment in West Bengal from outside the State became lesser than the previous decades. Position of the traditional industries like tea, jute and engineering were declining. This aggravated the unemployment situation in the State. Many industries became sick and were unable to compete with the industries of other States.

Emphasis on Public Sector

Establishment of new public sector units and nationalization of existing units were major goals of industrial policies till the implementation of New Economic Policy. The expansion of public sector units was initiated from the Second Five Year Plan. In 1951, there were only five central public undertakings while during 1996-97 the number reached to two hundred thirty six (Chatterjee, 2006). The approach of industrialization was mainly concentrated on heavy and basic industries and no major changes in industrial composition were observed at the end of 1980s. Almost all coal mines of the country came under public sector and a large number of sick textile units were nationalized from time to time. Both the policy resolutions of 1948 and 1956 had emphasized on public sector units of heavy industries with import substitution policy to stimulate the process of industrialization. But public sector units alone were incapable to improve the industrial scenario. The private sector units had no major scope for development. This huge number of sluggish running PSUs brought a crisis situation in Indian economy.

2.20 Role of GATT in international trade

General Agreement on Tariffs and Trade (GATT) was a legal agreement between many countries, whose overall purpose was to promote international trade by reducing or eliminating trade barriers such as tariffs or quotas. It set a multilateral trade agreements aimed at the abolition of quotas and the reduction of tariff duties among the contracting nations. It was formed in 1948 after World War II. GATT's primary purpose was to increase international trade by eliminating or reducing various tariffs, quotas and subsidies while maintaining meaningful regulations.

Role of GATT in Promoting International Trade

- Firstly, GATT established a set of standard to guide the contracting parties to participate in international trade practices. GATT stipulated several of basic principle to conduct the contracting parties in international business, such as General Most-Favored-Nation Treatment (Article II), Non-discriminatory Administration of Quantitative Restrictions (Article XIII), and General Elimination of Quantitative Regulations (Article XI) and so on in the "GATT 1947". Every contracting party should obey these basic principles when they were involved in trade relations, otherwise they would be condemned, even be taken revenge by other parties.

- Secondly, GATT reduced the tariff on the basis of mutual benefit, accelerate the trade liberalization after the World War II. GATT's major contribution was to reduce of tariffs by sponsoring "rounds" of multilateral negotiations.
- Thirdly, GATT reduced the discrimination in tariff and trade which promoted to reduce other trade barriers. As stated in the Article II: schedule of concession in "GATT 1947", "Each contracting party shall accord to the commerce of the other contracting parties' treatment no less favourable than that provided for in the appropriate Part of the appropriate Schedule annexed to this Agreement." According to this statement, GATT regulate the contracting parties cannot increase the levels of tariff as their wish, but some countries used other non-tariff barriers to promote their protectionism. Therefore, GATT claimed the contracting parties should not use other barriers to protect their own industries.
- Fourthly, GATT protected the benefits of the developing countries to a certain extent to international trade. One of the basic objectives of GATT was that "raising of standards of living and the progressive development of the economies of all contracting parties, and considering that the attainment of these objectives is particularly urgent for less-developed contracting parties."In order to achieve this objective, GATT established some special measures for less-developed countries, such as provide tariff protect for specific industries, quotas which are with the purpose of balance of payment.
- Finally, GATT acted as the "court of international trade", by providing a platform for contracting parties to negotiation and talk to settle disputes in international trade. One of the objectives of GATT was to settle the disputes between two or more parties. When two or more parties are involved in the international trade, it is inevitable that without disputes. Some of the disputes may be solved by the two parties themselves, however, some disputes could not be solved by themselves, without the help of the third party, and the disputes may be remaining unresolved for years. So it needed GATT to solve those disputes which could not solve by parties themselves.

2.21 Role of WTO in International Trade

WTO Agreements

How can you ensure that trade is as fair as possible, and as free as is practical? By negotiating rules and abiding by them. The WTO's rules – the agreements – are the result of negotiations between the members. The current set were the outcome of the 1986-94 Uruguay Round negotiations which included a major revision of the original General Agreement on Tariffs and Trade (GATT). GATT is now the WTO's principal rule-book for trade in goods. The Uruguay Round also created new rules for dealing with trade in services, relevant aspects of intellectual property, dispute settlement, and trade policy reviews. The complete set runs to some 30,000 pages consisting of about 30 agreements and separate commitments (called schedules) made by individual members in specific areas such as lower customs duty rates and services market-opening. Through these agreements, WTO members operate a non-discriminatory trading system that spells out their rights and their obligations. Each country receives guarantees that its exports will be treated fairly and consistently in other countries' markets. Each promises to do the same for imports into its own market. The system also gives developing countries some flexibility in implementing their commitments.

Goods

It all began with trade in goods. From 1947 to 1994, GATT was the forum for negotiating lower customs duty rates and other trade barriers; the text of the General Agreement spelt out important rules, particularly non-discrimination. Since 1995, the updated GATT has become the WTO's umbrella agreement for trade in goods. It has annexes dealing with specific sectors such as agriculture and textiles, and with specific issues such as state trading, product standards, subsidies and actions taken against dumping.

SERVICES

Banks, insurance firms, telecommunications companies, tour operators, hotel chains and transport companies looking to do business abroad can now enjoy the same principles of freer and fairer trade that originally only applied to trade in goods. These principles appear in the new General Agreement on Trade in Services (GATS). WTO members have also made individual commitments under GATS stating which of their services sectors they are willing to open to foreign competition, and how open those markets are.

The World Trade Organization (WTO) is the only international organization dealing with the global rules of trade between nations. Its main function is to ensure that trade flows as smoothly, predictably and freely as possible. It came into being in 1995. One of the youngest of the international organizations, the WTO is the successor to the General Agreement on Tariffs and Trade (GATT) established in the wake of the Second World War.

The WTO's overriding objective is to help trade flow smoothly, freely, fairly and predictably. The WTO has about 160 members, accounting for about 95% of world trade. Around 25 others are negotiating membership. Decisions are made by the entire membership. This is typically by consensus. A majority vote is also possible but it has never been used in the WTO, and was extremely rare under the WTO's predecessor, GATT. The WTO's agreements have been ratified in all members' parliaments. The WTO's top level decision-making body is the Ministerial Conference which meets at least once every two years.

Over three quarters of WTO members are developing or least-developed countries. All WTO agreements contain special provision for them, including longer time periods to implement agreements and commitments, measures to increase their trading opportunities and support to help them build the infrastructure for WTO work, handle disputes, and implement technical standards.

Objectives

(1) to set and enforce rules for international trade,
(2) to provide a forum for negotiating and monitoring further trade liberalization,
(3) to resolve trade disputes,
(4) to increase the transparency of decision-making processes

Functions of the World Trade Organisation

- It shall facilitate the implementation, administration and operation of the WTO trade agreements, such as multilateral trade agreements, plurilateral trade agreements.
- It shall provide forum for negotiations among its members concerning their multilateral trade relations.
- It shall administer the 'Understanding on Rules and Procedures' so as to handle trade disputes.
- It shall monitor national trade policies.
- It shall provide technical assistance and training for members of the developing countries.

- It shall cooperate with various international organisations like the IMF and the WB with the aim of achieving greater coherence in global economic policy-making.

Development and Trade

Over three-quarters of WTO members are developing or least developed countries. All WTO agreements contain special provision for them, including longer time periods to implement agreements and commitments, measures to increase their trading opportunities, provisions requiring all WTO members to safeguard their trade interests, and support to help them build the infrastructure for WTO work, handle disputes, and implement technical standards.

The 2001 Ministerial Conference in Doha set out tasks, including negotiations, for a wide range of issues concerning developing countries. Some people call the new negotiations the Doha Development Round. Before that, in 1997, a high-level meeting on trade initiatives and technical assistance for least-developed countries resulted in an "integrated framework" involving six intergovernmental agencies, to help least-developed countries increase their ability to trade, and some additional preferential market access agreements. A WTO Committee on Trade and Development, assisted by a Sub-Committee on Least-Developed Countries, looks at developing countries' special needs. Its responsibility includes implementation of the agreements, technical cooperation, and the increased participation of developing countries in the global trading system.

2.22 Issues Related to FDI in India's Retail Sector

Indian Retail Sector: An Overview

The presence of retail sector in India can be felt way back from the time of melas and haats, these were meant to cater the need of local people living nearby. The emergence of kiryana stores and mom & pop stores were observed after sometime. The Government also started supporting the retail sector and as a result a number of indigenous retail stores came into existence. With the passage of time, economy started becoming more open and a changed face of retail sector came into existence. The first few companies which came into existence to set up retail chains were from textile sector (S.Kumar's, Raymond etc). The Retail Industry can be broadly categorized as: Organized and Unorganized Retailing. Organized Retailing represents licensed retailers who got them registered for income tax, sales tax etc.

Unorganized Retailing represents retailers who are not registered for income tax, sales tax etc. and do not possess license for their workings, for example: street hawkers. Today India is fifth largest in the world in terms of retailing. The overall retail is projected to double to $1 trillion by 2020 from $600 billion in 2015, where modern retail is expected to grow 3 time to $180 billion by 2020 from $60 billion in 2015 as per BCG report on "Retail 2020: Retrospect, Reinvent, Rewrite". The retail sector is growing at a very faster pace and the key factors driving the growth are: Increasing no. of young population Double family income Increasing working women population Techno-savvy youngsters Nuclear families Rapid urbanization Customer liking towards modern shopping environment New retail formats with differentiated strategies Positive regulatory environment: promotion of ease of doing business concept through Make in India. As a result of ever changing demand of customers, liberalization of FDI policy, favourable responses of customers towards innovative products, the retail sector is able to attract big players to play in the field.

FDI Policy in India

According to Chawla et al (2016), 'FDI' means investment by non-resident entity/person resident outside India in the capital of an Indian company under Schedule 1 of Foreign Exchange Management (Transfer or Issue of Security by a Person Resident Outside India) Regulations, 2000.

In India, the Ministry of Commerce and Industry acts as a nodal agency for monitoring and reviewing FDI policy on a continuous basis. The FDI policy is notified through Press Notes released from time to time by Secretariat for Industrial Assistance (SIA), Department of Industrial Policy and Promotion (DIPP). The foreign investors are free to invest, except in few sectors, where prior approval from RBI or FIPB would be required. FDI Policy in Retail FDI policy related to Single Brand product Retail Trading FDI in Single Brand product retail trading is allowed 100% viz. automatic up to 49% and Government route beyond 49%. The circular shows that Foreign Investment in Single Brand retail trading aimed at attracting investments in production and marketing, improving the availability of such goods for the consumer, encouraging increased sourcing of goods from India, and enhancing competitiveness of Indian enterprises through access to global designs, technologies and management practices, as Per DIPP consolidated FDI policy circular of 2015. FDI in single brand product retail trading is subject to certain conditions such as:

Products to be sold should be of single brand only.

Products sold should be of same brand internationally.

'Single Brand' covers only products which are branded during manufacturing.

In case of proposals involving FDI beyond 51%, sourcing of 30% of the value of goods purchased will be done from India, preferably MSME's, village and cottage industries, artisans and craftsmen, in all sectors.

Applications would be processed firstly by DIPP and then by the FIPB for Government approval.

FDI policy related to Multi Brand Retail Trading

FDI in Multi Brand Retail Trading is allowed upto 51% through Government route, as per DIPP policy. FDI in Multi brand retail trading is subject to certain conditions such as:

Fresh agriculture produce (fruits, vegetables, flowers, grains, pulses, fresh poultry, fishery and meat products) may be unbranded.

The foreign investor must bring a minimum amount of US $ 100 million for investment.

At least 50% of the investment bought should be invested in 'back-end infrastructure' within three years. Expenditure on land cost and rentals will not be included in infrastructure development.

At least 30% of the products purchased must be sourced from Indian micro, small and medium industries(total investment in plant and machinery not exceeding US $ 2.0 million)

Government possess the first right of procurement on agriculture produce.

Retail outlets are allowed to be set up in cities with a population of more than 10 lakh as per 2011 census survey, or any other cities as per the decisions of the respective State Governments.

The policy for FDI is an enabling policy, the State Governments are set free for implementation of the policy.

Applications are to be processed firstly by DIPP followed by the FIPB for Government approval. Retail trading in any form by means of e-commerce is not permissible for FDI, engaged in multi brand or single brand retail trading.

2.23 Cashless Economy

Transformation into a cashless economy is an international issue and many of the countries already almost become cashless economy. A cashless economy is where financial transactions are not being done in the terms of currency notes, coins or physical cash money. It was in trend by barter age of cashless transaction and other methods of exchange like food crops or other goods (Humphrey, D.B, 2014).

However, the new concepts of cashless transactions in cashless economy are made with the help of digital currencies where legal tender (money) is exchanged and recorded only in the electronic digital form. So many challenges and opportunities are associated with the effects of digital transactions. Indian population where 98 per cent of total economic transactions by volume is being done through cash (Economic times, Nov. 23, 2016), much of the cash transactions being done in the country are small exchange for goods or services. The penetration of Pos terminals is not enough. Millions of people still do not have a bank account, internet network and connection is not proper, lack of knowledge to use online payment methods. These are some of the challenges are there in the country especially in small towns, rural areas and untapped markets in urban India, need to be resolved and people make assure that to adopt digitalisation in their payment system.

It is a big task in front of the government of India and their policy maker to transform their society into a cashless economy or less cash economy with the India's fast growing population. The introduction of cashless transaction has made the government of India to move towards cashless economy. India was the world's fastest growing major economy in the last quarter of 2014 (G20 an International Forum). India also topped the World Bank's growth outlook for 2015-16 for the first time with the economy rate having grown 7.6% in 2015- 16 and also expected to grow 8.0%+ in 2016-17. It is seen that growth of the Indian economy in the future is positive. India has already introduced some of the option of payment methods such as Ola money and PayTM accounts to pay rents via internet banking, Indian government has taken a decision of demonetisation (2016) by discontinuation of all 500 and 1000 banknotes, as it would no longer be recognised as legal tender. This move has been executed with the aim to curb the circulation of "black money" in the country and associated problems. India is going towards cashless economy very fast but it may be a long process for years to become complete cashless economy or less cash economy.

Challenges of Cashless Economy

In the process of digitalisations of an economy of the country, it is very important to assure that the availability of proper sources to setup require technology and sufficient manpower to provide prompt services in time. A bank account is a primary requirement for digitalisation. Hence Banks have a core responsibility to improve and develop them self -first. There are some challenges in the process of cashless economy in India.

a) Internet plays vital role between banks and customers to receive and deliver information, this forms of banking is describes as Internet banking (Reserve Bank of India, 2001).
b) Though bank accounts have been opened through PradhanMantri Jan-DhanYojna (PMJDY), most of them are not in operation. There were 256 million accounts, roughly one for every household. Opening accounts only does not empower citizens to make digital financial transactions. A recent investigation found that 23% of PMJDY accounts lie empty; 10

million accounts held only Rs. 1 and only 33% of all were ready to use their Rupay cards (September, 2016). 99% of households in both rural and urban India have at least one member with bank account.

In one of every five Indian households is a chief wage earner (CWE). CWE is a breadwinner who does not save money in a bank (MRSI 1998). We can see the Figure-7 and it is clear that still huge population are there in the country who are not operating their transaction through banks even having accounts in the bank.

c) India has a wide network of small retailers in all over countries area and most of they do not have enough resources to invest in electronic payment infrastructure to receive and make payments digitally.

d) The perception of consumers on use of credit and debit cards and belief that cash helps you negotiate better.

e) There is also vested interest in not moving towards cashless economy and most card and cash users fear that they will be charged more if they use cards. Further, non-users of credit cards are not aware of the benefits of credit cards due to lack of awareness of new technologies and financial literacy.

f) Indian banks are making it difficult for digital wallets issued by private sector companies to be used on the respective bank websites. It could be restrictions on using bank accounts to refill digital wallets or a lack of access to payments gateways. Regulators will have to take a tough stand against such rent-seeking behaviour by the banks.

Opportunities of Cashless Economy

The main advantage of cashless transactions is that a proper record of all economic transactions is possible to keep. It is remedy to remove black market or underground economies that often prove damaging national economies. Since, cash is the primary mode of transactions in money laundering and terrorism financing, a cashless economy discourages such activities. Central government also get benefit from such cashless transactions as it allows central control of money supply and it is easier to monitor income tax paid by an individual. Cashless transactions are helpful in the context of negative global inflation and quantitative easing. Going cashless is also reducing the levels of corruption prevalent in the country.

Some more benefits of cashless economy are as follows:

a) Real estate prices will reduce because of curbs on black money as most of black money is invested in Real estate.
b) There is a process of issuance of currency where government bear costs against designing, developing, printing, storing, transporting and placing etc. All this can be avoided by digitalisation of cash transactions. In Financial year 2015, RBI spent Rs. 27 billion for currency issuance and management.
c) It pave way for universal availability of banking services to all as no physical infrastructure is needed other than digital. People can make their payments and receive globally.
d) There is greater efficiency in welfare programmes as money is wired directly into the accounts of recipients with great transparency.
e) There are efficiency gains as transaction costs will also come down by using methods of digital payments.
f) In India, 1 in 7 notes is supposed to be fake, which has a huge negative impact on economy. By going cashless transaction, accountability of notes and coins in circulation will be possible.
g) Soiled or tobacco stained notes full of germs are a norm in India. There are many such incidents in our life where we knowingly or unknowing give and take germs in the form of rupee notes. This can be avoided.

h) In the recent trends of digitalisation will reduce costs of operating ATM's. i) Impact of digital transactions increase in GDP by emerging markets resulting growth of country economy.

Self-Assessment Tests

1) Write the positive and negative impacts of green revolution in India.
2) How was the milk revolution of India attained?
3) Write the differences between SEZ and EEZ.
4) What do you mean by digital divide?
5) Bring out the concept of e-commerce. How does it differ from traditional commerce?
6) What do you mean by liberalisation of economy? Write the effects of liberalisation policy on Indian economy.
7) Write in short the different industrial policies adopted in different plan periods.
8) How would you measure agricultural productivity?

Selected Readings

Aggarwal, A. (2004). *Export Processing Zones in India: Analysis of the Export Performance*, Indian Council for Research on International Economic Relations, Working Paper No. 148.

Atkins, P. J. (1989). Operation Flood, Geography, 74, Pp. 259-262.

Boschma, R. and R. Martin (2010).The Aims and Scope of Evolutionary Economic Geography.Urban & Regional Research Centre, Utrecht University.

Candler, W. and N. Kumar (1998).*India: The Dairy Revolution*, The World Bank, Washington, D.C.

Chawla, J., R. Agrawal& B. Sharma (2016). FDI in Indian Retail Sector: A Review. International Journal of BRIC Business Research (IJBBR), 5(2).

Dutta, S. (2012).*Assessment of Agricultural Efficiency and Productivity: A Study of Hugli District, West Bengal, India*. International Journal of Current Research, 4(11), p. 190-5.

Firth, J. & E. Liu (2018).Manufacturing Underdevelopment:India's Freight Equalization Scheme, and the Long-runEffects of Distortions on the Geography of Production. Available at http://barrett.dyson.cornell.edu/NEUDC/paper_316.pdf.

Kumar, A., K. Kumar, N. Kaushik, S. Sharma and S. Mishra (2010).*Renewable energy in India: Current status and future potentials*. Renewable and Sustainable Energy Reviews, 14,2434–2442.

Kurien, V. (2000).*India's Milk Revolution*, Bulletin of the Nutrition Foundation of India, 21 (1).

Maddison, A. (2007). The world economy volume 1: A millennial perspective volume 2:Historical statistics. Academic Foundation.

Mittal, S. (2018).*Positive impact of privatization on Indian economy*. International Journal of Advance Research and Development, 3(7), p. 110-3.

Papadopoulos, N. & S. Malhotra (2007). Export Processing Zones in Development and International Marketing: An Integrative Review and Research Agenda, Journal of Macromarketing, 27 (2), https://doi.org/10.1177/0276146707300070.

Patuzi, D. (2015). *The concept of the Economic Exclusive Zone*. Academic Journal of Business, Administration, Law and Social Sciences, 1(1), 149–159.

Phillipson, N. (2010). Adam Smith: an enlightened life. Penguin UK.

Sisaye, S. & E. Stommes (1985). *Green revolution as a planned intervention strategy for agricultural development: A systems perspective*, Public Administration and Development, January/March, Pp. 39-55, https://doi.org/10.1002/pad.4230050105.

Sebby, K. (2010). *The Green Revolution of The 1960's and Its Impact on Small Farmers in India*.Undergraduate Thesis.University of Nebraska.

Shiva, V. (1993).The Violence of the Green Revolution: Third World Agriculture, Ecology and Politics. 2nd ed. London, UK: Zed Books Ltd.

Singhraul, B. P. & Y. S. Garwal (2018).Cashless Economy – Challenges and Opportunities in India.Pacific Business Review International, 10 (9), p. 54-63.

Somani, A., G. Sultania, A. Kumar & S. Doshi (2013).Rural Development Plans in India.Technical Paper, November, Indian Institute of Technology, Bombay.

Srinuan, C. & E. Bohlin (2011). Understanding the Digital Divide: A Literature Survey and ways Forward. 22nd European Regional Conference of the InternationalTelecommunications Society (ITS2011), Budapest, 18 - 21 September, 2011: Innovative ICTApplications - Emerging Regulatory, Economic and Policy Issues

Starr, P. (1988). *The Meaning of Privatization*. Yale Law and Policy Review, 6(1), p. 6-41.

Whittlesey, D. (1936). *Major Agricultural Regions of the Earth*. Annals of the Association of American Geographers, 26 (4), Pp. 199-240.

Webliography

1) http://shodhganga.inflibnet.ac.in/bitstream/10603/138279/14/14_chapter%206.pdf
2) http://www.mikeigbokwe.com/new1/WTO%20and%20its%20role%20in%20the%20international%20trade..pdf
3) http://www.yourarticlelibrary.com/geography/economic-geography/economic-geography-definition-scope-and-importance/74514
4) https://en.wikipedia.org/wiki/Category:Ring_roads_in_India
5) https://en.wikipedia.org/wiki/Digital_divide
6) https://en.wikipedia.org/wiki/Economic_geography
7) https://en.wikipedia.org/wiki/Exclusive_economic_zone
8) https://en.wikipedia.org/wiki/Golden_Quadrilateral
9) https://en.wikipedia.org/wiki/Green_Revolution_in_India
10) https://en.wikipedia.org/wiki/List_of_ring_roads
11) https://en.wikipedia.org/wiki/Special_economic_zone
12) https://www.brookings.edu/wp-content/uploads/2016/07/selfenforcingtrade_chapter.pdf
13) https://www.lawteacher.net/free-law-essays/international-law/role-of-gatt-in-promoting-international-trade-international-law-essay.php
14) https://www.sourcewatch.org/index.php/The_Green_Revolution_in_India
15) Online Exim Encyclopaedia
16) http://www.yourarticlelibrary.com/geography/resources/resource-meaning-concept-and-its-classification-geography/74529

17) https://www.oecd-ilibrary.org/docserver/9789264185715-17-en.pdf?expires=1566895291&id=id&accname=guest&checksum=8D306EA7AA33745DC045A127D549E684
18) https://www.worldatlas.com/articles/what-is-a-flow-resource.html
19) https://www.nrel.gov/docs/fy01osti/27955.pdf
20) http://www.iiasa.ac.at/web/home/research/Flagship-Projects/Global-Energy-Assessment/GEA_Chapter11_renewables_hires.pdf
21) https://ask.learncbse.in/t/classify-resources-on-the-basis-of-ownership-with-examples/23341
22) https://shodhganga.inflibnet.ac.in/bitstream/10603/49943/6/06_chapter%201.pdf
23) http://www.commonlii.org/in/journals/NALSARLawRw/2013/8.pdf
24) https://www.investopedia.com/terms/f/forwardcontract.asp
25) https://www.tutorialspoint.com/e_commerce/e_commerce_tutorial.pdf
26) https://pdfs.semanticscholar.org/173c/819ce3fccafdc4f64af71fd6868e815580ad.pdf
27) https://people.dsv.su.se/~miwebb/kapitel/kap38.pdf
28) https://www.economicshelp.org/blog/58802/trade/the-importance-of-international-trade/
29) https://www.quora.com/What-is-the-scope-and-the-nature-of-the-economic-geography
30) http://www.yourarticlelibrary.com/geography/economic-geography-scope-of-economic-geography-useful-notes/25430
31) http://www.yourarticlelibrary.com/geography/economic-geography/economic-geography-definition-scope-and-importance/74514
32) http://maryannfeldman.web.unc.edu/files/2011/11/Economic_Growth_and_Economic_Development_2016_Feldman_Storper.pdf
33) http://www.yourarticlelibrary.com/geography/whittleseys-classification-of-agricultural-regions/42221
34) http://www.yourarticlelibrary.com/agriculture/measurement-of-agricultural-productivity-and-efficiency/44426
35) http://policydialogue.org/files/events/Fujita_green_rev_in_india.pdf
36) http://indairyasso.org/pdf/info/Dr-Kurien.pdf
37) https://fairgaze.com/FGNews/white-revolution-and-its-impacts-on-indian-economy_71639.html
38) http://ncert.nic.in/ncerts/l/hess405.pdf
39) http://www.yourarticlelibrary.com/industries/industrial-regions-8-major-industrial-regions-of-india/14159
40) https://steel.gov.in/sites/default/files/industrialPolicy.pdf
41) https://shodhganga.inflibnet.ac.in/bitstream/10603/59577/9/09_chapter%203.pdf
42) https://www.jbs.cam.ac.uk/fileadmin/user_upload/research/centres/india/downloads/research_kattuman.pdf
43) http://dspace.vpmthane.org:8080/jspui/bitstream/123456789/4186/1/FC%20Sem%202%20CONCEPT%20OF%20LIBERALIZATION%20%20PRIVATIZATION%20AND%20GLOBALIZATION.pdf
44) https://shodhganga.inflibnet.ac.in/bitstream/10603/9795/9/09_chapter%204.pdf
45) https://shodhganga.inflibnet.ac.in/bitstream/10603/163879/18/18_summary%20critical%20appreciation%20of%20the%20policies%20and%20conclusion.pdf
46) https://www.wto.org/english/res_e/doload_e/inbr_e.pdf

Disclaimer

This self-learning material is based on different books, journals and web-sources.

Self-Learning Material (SLM) for the Course of

M.A. / M.Sc. in Geography

Directorate of Open and Distance Learning (DODL)

University of Kalyani

Paper: GEO310T

(Total Credit – 4; Total Marks – 100: Internal Evaluation – 20 + Semester-end Examination - 80)

University of Kalyani

Kalyani

Nadia – 741235

West Bengal

Table: Course Material Compilation Team

Sl. No.	Names	Designation and Affiliation	Group of the SLM
1	Dr. Sayan Choudhary	Assistant Professor (Contractual) of Geography, Directorate of Open and Distance Learning (DODL), University of Kalyani	GEO310T.1: Advanced Cartography and Geoinformatics
2	Dr. Bhaswati Mondal	Assistant Professor (Contractual) of Geography, Directorate of Open and Distance Learning (DODL), University of Kalyani	GEO310T.2: Research Methodology

December 2019

Directorate of Open and Distance Learning, University of Kalyani

Published by the Directorate of Open and Distance Learning,

University of Kalyani, Kalyani-741235, West Bengal.

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2017 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor (Dr.) Sankar Kumar Ghosh, Hon'ble Vice- Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks is also due to the Course Writers-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and co-ordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self writing and collected from e-book, journals and websites.

Director

Directorate of Open and Distance Learning

University of Kalyani

SEMESTER-III

Paper – GEO310T: (Total Credit - 4, Total Marks – 100)

GROUP – GEO310T.1:ADVANCED CARTOGRAPHY AND GEOINFORMATICS

(Credit – 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit 1. Cartography – nature, scope and development;

Unit 2. Nature and types of Geoid; Concept of Datum with special reference to NAD, Everest and WGS-84; Principles of Spherical Trigonometry

Unit 3. Principles and properties of UTM Projections

Unit 4. Concept of Geoinformatics; Remote Sensing Platforms and Sensors;

Unit 5. Nature of EMR, EMS, and interaction with atmosphere and surface materials;

Unit 6. Different sensors and resolutions;

Unit 7. Digital Image Processing – principles and approaches;

Unit 8: Analytical Modelling in GIS, GPS-GIS integration

Mode of Internal Evaluation:

GROUP – GEO310T.2: RESEARCH METHODOLOGY

(Credit – 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit-9: Spectrum of Geographical Research and its approaches: Inductive and Deductive, Objective and Subjective

Unit-10: Critical issues in major areas of geographical research;

Unit-11: Identification of Research Problem

Unit-12: Hypothesis Building

Unit-13: Methods of Sampling and sample design

Unit-14: Methodological orientation: Quantitative and Qualitative

Unit-15: Abstract and summery and synopsis: their differences

Unit-16: Referencing style and preparation of Bibliography

Mode of Internal Evaluation:

Contents

GROUP – GEO310T.1:ADVANCED CARTOGRAPHY AND GEOINFORMATICS

(Credit – 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

1.1 INTRODUCTION

Cartography – the science and art of map making, has evolved through different phases of development in science and technology. Remote Sensing and GIS have revolutionised the map making industry. The present section discusses about the different sensors and their resolution and their applications. Digital image processing – their principle sand approaches have been discussed. Finally, this chapter introduces analytical modelling in GIS, and GPS –GIS integration.

1.2 LEARNING OBJECTIVES

The present section aims to introduce the following topics -

- Cartography – nature, scope and development;
- Nature and types of Geoid; Concept of Datum with special reference to NAD, Everest and WGS-84; Principles of Spherical Trigonometry
- Principles and properties of UTM Projections
- Concept of Geoinformatics; Remote Sensing Platforms and Sensors;
- Nature of EMR, EMS, and interaction with atmosphere and surface materials;
- Different sensors and resolutions;
- Digital Image Processing – principles and approaches;
- Analytical Modelling in GIS, GPS-GIS integration

1.3 ASSESSMENT OF PRIOR KNOWLEDGE

Discussion about principles of maps and map projections are necessary. Discussion about latitude and longitude is necessary. Familiarity with satellite imagery and aerial photographs are necessary.

1.4 LEARNING ACTIVITIES

Identification of different features from maps and imagery may be done.

1.5 FEEDBACK OF LEARNING ACTIVITIES

Discussion and debate on different application of remotely sensed products in class room will helpful.

1.6 EXAMPLES AND ILLUSTRATIONS

UNIT 1: CARTOGRAPHY – NATURE, SCOPE AND DEVELOPMENT

Cartography is the study and practice of making maps. Combining science, aesthetics, and technique, cartography builds on the premise that reality can be modelled in ways that communicate spatial information effectively.

The fundamental problems of traditional cartography are to:

- Set the map's agenda and select traits of the object to be mapped. This is the concern of map editing. Traits may be physical, such as roads or land masses, or may be abstract, such as toponyms or political boundaries.
- Represent the terrain of the mapped object on flat media. This is the concern of map projections.
- Eliminate characteristics of the mapped object that are not relevant to the map's purpose. This is the concern of generalization.
- Reduce the complexity of the characteristics that will be mapped. This is also the concern of generalization.
- Orchestrate the elements of the map to best convey its message to its audience. This is the concern of map design.

Modern cartography constitutes many theoretical and practical foundations of geographic information systems.

History

Ancient times

What is the earliest known map is a matter of some debate, both because the term "map" is not well-defined and because some artifacts that might be maps might actually be something else. A wall painting that might depict the ancient Anatolian city of Çatalhöyük (previously known as CatalHuyuk or ÇatalHüyük) has been dated to the late 7th millennium BCE. Among the prehistoric alpine rock carvings of Mount Bego (France) and Valcamonica (Italy), dated to the 4th millennium BCE, geometric patterns consisting of dotted rectangles and lines are widely interpreted in archaeological literature as a depiction of cultivated plots. Other known maps of the ancient world include the Minoan "House of the Admiral" wall painting from c. 1600 BCE, showing a seaside community in an oblique perspective, and an engraved map of the holy Babylonian city of Nippur, from the Kassite period (14th – 12th centuries BCE). The oldest surviving world maps are from 9th century BCE Babylonia. One shows Babylon on the Euphrates, surrounded by Assyria, Urartu and several cities, all, in turn, surrounded by a "bitter river" (Oceanus). Another depicts Babylon as being north of the center of the world.

The ancient Greeks and Romans created maps from the time of Anaximander in the 6th century BCE. In the 2nd century CE, Ptolemy wrote his treatise on cartography, Geographia. This contained Ptolemy's world map – the world then known to Western society *(Ecumene)*. As early as the 8th century, Arab scholars were translating the works of the Greek geographers into Arabic.

In ancient China, geographical literature dates to the 5th century BCE. The oldest extant Chinese maps come from the State of Qin, dated back to the 4th century BCE, during the Warring States period. In the book of the *Xin Yi Xiang Fa Yao*, published in 1092 by the Chinese scientist Su Song, a star map on the equidistant cylindrical projection. Although this method of charting seems to have existed in China even before this publication and scientist, the greatest significance of the star maps by Su Song is that they represent the oldest existent star maps in printed form.

Early forms of cartography of India included depictions of the pole star and surrounding constellations. These charts may have been used for navigation.

Middle Ages and Renaissance

Mappae mundi ("maps of the world") are the medieval European maps of the world. About 1,100 of these are known to have survived: of these, some 900 are found illustrating manuscripts and the remainder exist as stand-alone documents.

The Arab geographer Muhammad al-Idrisi produced his medieval atlas *Tabula Rogeriana (Book of Roger)* in 1154. By combining the knowledge of Africa, the Indian Ocean, Europe, and the Far East (which he learned through contemporary accounts from Arab merchants and explorers) with the information he inherited from the classical geographers, he was able to write detailed descriptions of a multitude of countries. Along with the substantial text he had written, he created a world map influenced mostly by the Ptolemaic conception of the world, but with significant influence from multiple Arab geographers. It remained the most accurate world map for the next three centuries. The map was divided into seven climatic zones, with detailed descriptions of each zone. As part of this work, a smaller, circular map was made depicting the south on top and Arabia in the center. Al-Idrisi also made an estimate of the circumference of the world, accurate to within 10%.

In the Age of Exploration, from the 15th century to the 17th century, European cartographers both copied earlier maps (some of which had been passed down for centuries) and drew their own, based on explorers' observations and new surveying techniques. The invention of the magnetic compass, telescope and sextant enabled increasing accuracy. In 1492, Martin Behaim, a German cartographer, made the oldest extant globe of the Earth.

In 1507, Martin Waldseemüller produced a globular world map and a large 12-panel world wall map (*UniversalisCosmographia*) bearing the first use of the name "America". Portuguese cartographer Diego Ribero was the author of the first known planisphere with a graduated Equator (1527). Italian cartographer Battista Agnese produced at least 71 manuscript atlases of sea charts. Johannes Werner refined and promoted the Werner projection. This was an equal-area, heart-shaped world map projection (generally called a cordiform projection) which was used in the 16th and 17th centuries. Over time, other iterations of this map type arose; most notable are the sinusoidal projection and the Bonne projection. The Werner projection places its standard parallel at the North Pole; a sinusoidal projection places its standard parallel at the equator; and the Bonne projection is intermediate between the two.

In 1569, mapmaker Gerardus Mercator first published a map based on his Mercator projection, which uses equally-spaced parallel vertical lines of longitude and parallel latitude lines spaced farther apart as they get farther away from the equator. By this construction, courses of constant bearing are conveniently represented as straight lines for navigation. The same property limits its value as a general-purpose world map because regions are shown as increasingly larger than they actually are the further from the equator they are. Mercator is also credited as the first to use the word "atlas" to describe a collection of maps. In the later years of his life, Mercator resolved to create his Atlas, a book filled with many maps of different regions of the world, as well as a chronological history of the world from the Earth's creation by God until 1568. He was unable to complete it to his satisfaction before he died. Still, some additions were made to the Atlas after his death and new editions were published after his death.

In the Renaissance, maps were used to impress viewers and establish the owner's reputation as sophisticated, educated, and worldly. Because of this, towards the end of the Renaissance, maps were displayed with equal importance of painting, sculptures, and other pieces of art. In the sixteenth century, maps were becoming increasingly available to consumers through the introduction of printmaking, with about 10% of Venetian homes having some sort of map by the late 1500s.

There were three main functions of maps in the Renaissance:

- General descriptions of the world
- Navigation and wayfinding
- Land surveying and property management

In medieval times, written directions of how to get somewhere were more common than the use of maps. With the Renaissance, cartography began to be seen as a metaphor for power. Political leaders could lay claim on territories through the use of maps and this was greatly aided by the religious and colonial expansion of Europe. The most commonly mapped places during the Renaissance were the Holy Land and other religious places.

In the late 1400s to the late 1500s, Rome, Florence, and Venice dominated map making and trade. It started in Florence in the mid to late 1400s. Map trade quickly shifted to Rome and Venice but then was overtaken by atlas makers in the late 16th century. Map publishing in Venice was completed with humanities and book publishing in mind, rather than just informational use.

Printing technology

There were two main printmaking technologies in the Renaissance: woodcut and copper-plate intaglio, referring to the medium used to transfer the image onto paper.

In woodcut, the map image is created as a relief chiseled from medium-grain hardwood. The areas intended to be printed are inked and pressed against the sheet. Being raised from the rest of the block, the map lines cause indentations in the paper that can often be felt on the back of the map. There are advantages to using relief to make maps. For one, a printmaker doesn't need a press because the maps could be developed as rubbings. Woodblock is durable enough to be used many times before defects appear. Existing printing presses can be used to create the prints rather than having to create a new one. On the other hand, it is hard to achieve fine detail with the relief technique. Inconsistencies in linework are more apparent in woodcut than in intaglio. To improve quality in the late fifteenth century, a style of relief craftsmanship developed using fine chisels to carve the wood, rather than the more commonly used knife.

In intaglio, lines are engraved into workable metals, typically copper but sometimes brass. The engraver spreads a thin sheet of wax over the metal plate and uses ink to draw the details. Then, the engraver traces the lines with a stylus to etch them into the plate beneath. The engraver can also use styli to prick holes along the drawn lines, trace along them with colored chalk, and then engrave the map. Lines going in the same direction are carved at the same time, and then the plate is turned to carve lines going in a different direction. To print from the finished plate, ink is spread over the metal surface and scraped off such that it remains only in the etched channels. Then the plate is pressed forcibly against the paper so that the ink in the channels is transferred to the paper. The pressing is so forceful that it leaves a "plate mark" around the border of the map at the edge of the plate, within which the paper is depressed compared to the margins. Copper and other metals were expensive at the time, so the plate was often reused for new maps or melted down for other purposes.

Whether woodcut or intaglio, the printed map is hung out to dry. Once dry, it is usually placed in another press of flatten the paper. Any type of paper that was available at the time could be used to print the map on, but thicker paper was more durable.

Both relief and intaglio were used about equally by the end of the fifteenth century.

Lettering

Lettering in mapmaking is important for denoting information. Fine lettering is difficult in woodcut, where it often turned out square and blocky, contrary to the stylized, rounded writing style popular in Italy at the time. To improve quality, mapmakers developed fine chisels to carve the relief. Intaglio lettering did not suffer the troubles of a coarse medium and so was able to express the looping cursive that came to be known as cancellaresca. There were custom-made reverse punches that were also used in metal engraving alongside freehand lettering.

Colour

The first use of colour in map making cannot be narrowed down to one reason. There are arguments that colour started as a way to indicate information on the map, with aesthetics coming second. There are also arguments that colour was first used on maps for aesthetics but then evolved into conveying information. Either way, many maps of the Renaissance left the publisher without being coloured, a practice that continued all the way into the 1800s. However, most publishers accepted orders from their patrons to have their maps or atlases coloured if they wished. Because all colouring was done by hand, the patron could request simple, cheap colour, or more expensive, elaborate colour, even going so far as silver or gold gilding. The simplest colouring was merely outlines, such as of borders and along rivers. Wash colour meant painting regions with inks or watercolours. Limning meant adding silver and gold leaf to the map to illuminate lettering, heraldic arms, or other decorative elements.

The Enlightenment

Maps of the Enlightenment period practically universally used copper plate intaglio, having abandoned the fragile, coarse woodcut technology. Use of map projections evolved, with the double hemisphere being very common and Mercator's prestigious navigational projection gradually making more appearances.

Due to the paucity of information and the immense difficulty of surveying during the period, mapmakers frequently plagiarized material without giving credit to the original cartographer. For example, a famous map of North America known as the "Beaver Map" was published in 1715 by Herman Moll. This map is a close reproduction of a 1698 work by Nicolas de Fer. De Fer, in turn, had copied images that were first printed in books by Louis Hennepin, published in 1697, and François Du Creux, in 1664. By the late 18th century, mapmakers often credited the original publisher with something along the lines of, "After [the original cartographer]" in the map's title or cartouche.

Modern period

In cartography, technology has continually changed in order to meet the demands of new generations of mapmakers and map users. The first maps were produced manually, with brushes and parchment; so they varied in quality and were limited in distribution. The advent of magnetic devices, such as the compass and much later, magnetic storage devices, allowed for the creation of far more accurate maps and the ability to store and manipulate them digitally.

Advances in mechanical devices such as the printing press, quadrant and vernier, allowed the mass production of maps and the creation of accurate reproductions from more accurate data. Hartmann Schedel was one of the first cartographers to use the printing press to make maps more widely available. Optical technology, such as the telescope, sextant and other devices that use telescopes, allowed accurate land surveys and allowed mapmakers and navigators to find their latitude by measuring angles to the North Star at night or the Sun at noon.

Advances in photochemical technology, such as the lithographic and photochemical processes, make possible maps with fine details, which do not distort in shape and which resist moisture and wear. This also eliminated the need for engraving, which further speeded up map production.

In the 20th century, aerial photography, satellite imagery, and remote sensing provided efficient, precise methods for mapping physical features, such as coastlines, roads, buildings, watersheds, and topography. The United States Geological Survey has devised multiple new map projections, notably the Space Oblique Mercator for interpreting satellite ground tracks for mapping the surface. The use of satellites and space telescopes now allows researchers to map other planets and moons in outer space. Advances in electronic technology ushered in another revolution in cartography: ready availability of computers and peripherals such as monitors, plotters, printers, scanners (remote and document) and analytic stereo plotters, along with computer programs for visualization, image processing, spatial analysis, and database management, democratized and greatly expanded the making of maps. The ability to superimpose spatially located variables onto existing maps created new uses for maps and new industries to explore and exploit these potentials. See also digital raster graphic.

These days most commercial-quality maps are made using software of three main types: CAD, GIS and specialized illustration software. Spatial information can be stored in a database, from which it can be extracted on demand. These tools lead to increasingly dynamic, interactive maps that can be manipulated digitally.

Field-rugged computers, GPS, and laser rangefinders make it possible to create maps directly from measurements made on site.

Deconstruction

There are technical and cultural aspects to producing maps. In this sense, maps can sometimes be said to be biased. The study of bias, influence, and agenda in making a map is what comprise a map's deconstruction. A central tenet of deconstructionism is that maps have power. Other assertions are that maps are inherently biased and that we search for metaphor and rhetoric in maps.

It is claimed that the Europeans promoted an "epistemological" understanding of the map as early as the 17th century. An example of this understanding is that "[European reproduction of terrain on maps] reality can be expressed in mathematical terms; that systematic observation and measurement offer the only route to cartographic truth…". 17th-century map-makers were careful and precise in their strategic approaches to maps based on a scientific model of knowledge. Popular belief at the time was that this scientific approach to cartography was immune to the social atmosphere.

A common belief is that science heads in a direction of progress, and thus leads to more accurate representations of maps. In this belief European maps must be superior to others, which necessarily employed different map-making skills. "There was a 'not cartography' land where lurked an army of inaccurate, heretical, subjective, valuative, and ideologically distorted images. Cartographers developed a 'sense of the other' in relation to nonconforming maps."

Although cartography has been a target of much criticism in recent decades, a cartographer's 'black box' always seemed to be naturally defended to the point where it overcame the criticism. However, to later scholars in the field, it was evident that cultural influences dominate map-making. For instance, certain abstracts on maps and the map-making society itself describe the social influences on the production of maps. This social play on cartographic knowledge "…produces the 'order' of [maps'] features and the 'hierarchies of its practices.'"

Depictions of Africa are a common target of deconstructionism. According to deconstructionist models, cartography was used for strategic purposes associated with imperialism and as instruments and representations of power during the conquest of Africa. The depiction of Africa and the low latitudes in general on the Mercator projection has been interpreted as imperialistic and as symbolic of subjugation due to the diminished proportions of those regions compared to higher latitudes where the European powers were concentrated.

Maps furthered imperialism and colonization of Africa in practical ways by showing basic information like roads, terrain, natural resources, settlements, and communities. Through this, maps made European commerce in Africa possible by showing potential commercial routes and made natural resource extraction possible by depicting locations of resources. Such maps also enabled military conquests and made them more efficient, and imperial nations further used them to put their conquests on display. These same maps were then used to cement territorial claims, such as at the Berlin Conference of 1884–1885.

Before 1749, maps of the African continent had African kingdoms drawn with assumed or contrived boundaries, with unknown or unexplored areas having drawings of animals, imaginary physical geographic features, and descriptive texts. In 1748, Jean B. B. d'Anville created the first map of the African continent that had blank spaces to represent the unknown territory. This was revolutionary in cartography and the representation of power associated with map making.

Map types

General vs. thematic cartography

In understanding basic maps, the field of cartography can be divided into two general categories: general cartography and thematic cartography. General cartography involves those maps that are constructed for a general audience and thus contain a variety of features. General maps exhibit many reference and location systems and often are produced in a series. For example, the 1:24,000 scale topographic maps of the United States Geological Survey (USGS) are a standard as compared to the 1:50,000 scale Canadian maps. The government of the UK produces the classic 1:50,000 (replacing the older 1 inch to 1 mile) "Ordnance Survey" maps of the entire UK and with a range of correlated larger- and smaller-scale maps of great detail. Many private mapping companies have also produced thematic map series.

Thematic cartography involves maps of specific geographic themes, oriented toward specific audiences. A couple of examples might be a dot map showing corn production in Indiana or a shaded area map of Ohio counties, divided into numerical choropleth classes. As the volume of geographic data has exploded over the last century, thematic cartography has become increasingly useful and necessary to interpret spatial, cultural and social data.

A third type of map is known as an "orienteering," or special purpose map. This type of map falls somewhere between thematic and general maps. They combine general map elements with thematic attributes in order to design a map with a specific audience in mind. Oftentimes, the type of audience an orienteering map is made for is in a particular industry or occupation. An example of this kind of map would be a municipal utility map.

Topographic vs. topological

A topographic map is primarily concerned with the topographic description of a place, including (especially in the 20th and 21st centuries) the use of contour lines showing elevation. Terrain or relief can be shown in a variety of ways (see Cartographic relief depiction). In the present era, one of the most widespread and advanced methods used to form topographic maps is to use computer software to generate digital elevation models which show shaded relief. Before such software existed, cartographers had to draw shaded relief by hand. One cartographer who is respected as a master of hand-drawn shaded relief is the Swiss professor Eduard Imhof whose efforts in hill shading were so influential that his method became used around the world despite it being so labor-intensive.

A topological map is a very general type of map, the kind one might sketch on a napkin. It often disregards scale and detail in the interest of clarity of communicating specific route or relational information. Beck's London Underground map is an iconic example. Although the most widely used map of "The Tube," it preserves little of reality: it varies scale constantly and abruptly, it straightens curved tracks, and it contorts directions. The only topography on it is the River Thames, letting the reader know whether a station is north or south of the river. That and the topology of station order and interchanges between train lines are all that is left of the geographic space. Yet those are all a typical passenger wishes to know, so the map fulfills its purpose.

Map design

Map purpose and selection of information

Arthur H. Robinson, an American cartographer influential in thematic cartography, stated that a map not properly designed "will be a cartographic failure." He also claimed, when considering all aspects of cartography, that "map design is perhaps the most complex." Robinson codified the mapmaker's understanding that a map must be designed foremost with consideration to the audience and its needs.

From the very beginning of mapmaking, maps "have been made for some particular purpose or set of purposes". The intent of the map should be illustrated in a manner in which the percipient acknowledges its purpose in a timely fashion. The term *percipient* refers to the person receiving information and was coined by Robinson. The principle of figure-ground refers to this notion of engaging the user by presenting a clear presentation, leaving no confusion concerning the purpose of the map. This will enhance the user's experience and keep their attention. If the user is unable to identify what is being demonstrated in a reasonable fashion, the map may be regarded as useless.

Making a meaningful map is the ultimate goal. Alan MacEachren explains that a well designed map "is convincing because it implies authenticity". An interesting map will no doubt engage a reader. Information richness or a map that is multivariate shows relationships within the map. Showing several variables allows comparison, which adds to the meaningfulness of the map. This also generates hypothesis and stimulates ideas and perhaps further research. In order to convey the message of the map, the creator must design it in a manner which will aid the reader in the overall understanding of its purpose. The title of a map may provide the "needed link" necessary for communicating that message, but the overall design of the map fosters the manner in which the reader interprets it.

In the 21st century it is possible to find a map of virtually anything from the inner workings of the human body to the virtual worlds of cyberspace. Therefore, there are now a huge variety of different styles and types of map – for example, one area which has evolved a specific and recognisable variation are those used by public transport organisations to guide passengers, namely urban rail and metro maps, many of which are loosely based on 45 degree angles as originally perfected by Harry Beck and George Dow.

Naming conventions

Most maps use text to label places and for such things as the map title, legend and other information. Although maps are often made in one specific language, place names often differ between languages. So a map made in English may use the name *Germany* for that country, while a German map would use *Deutschland* and a French map *Allemagne*. A non-native term for a place is referred to as an exonym.

In some cases the correct name is not clear. For example, the nation of Burma officially changed its name to Myanmar, but many nations do not recognize the ruling junta and continue to use *Burma*. Sometimes an official name change is resisted in other languages and the older name may remain in common use. Examples include the use of *Saigon* for Ho Chi Minh City, *Bangkok* for Krung Thep and *Ivory Coast* for Côte d'Ivoire.

Difficulties arise when transliteration or transcription between writing systems is required. Some well-known places have well-established names in other languages and writing systems, such as *Russia* or *Rußland* for Росси́я, but in other cases a system of transliteration or transcription is required. Even in the former case, the exclusive use of an exonym may be unhelpful for the map user. It will not be much use for an English user of a map of Italy to show Livorno *only* as "Leghorn" when road signs and railway timetables show it as "Livorno". In transliteration, the characters in one script are represented by characters in another. For example, the Cyrillic letter *Р* is usually written as *R* in the Latin script, although in many cases it is not as simple as a one-for-one equivalence. Systems exist for transliteration of Arabic, but the results may vary. For example, the Yemeni city of Mocha is written variously in English as Mocha, Al Mukha, al-Mukhā, Mocca and Moka. Transliteration systems are based on relating written symbols to one another, while transcription is the attempt to spell in one language the phonetic sounds of another. Chinese writing is now usually converted to the Latin alphabet through the Pinyin phonetic transcription systems. Other systems were used in the past, such as Wade-Giles, resulting in the city being spelled *Beijing* on newer English maps and *Peking* on older ones.

Further difficulties arise when countries, especially former colonies, do not have a strong national geographic naming standard. In such cases, cartographers may have to choose between various phonetic spellings of local names versus older imposed, sometimes resented, colonial names. Some countries have multiple official languages, resulting in multiple official placenames. For example, the capital of Belgium is both *Brussel* and *Bruxelles*. In Canada, English and French are official languages and places have names in both languages. British Columbia is also officially named *la Colombie-Britannique*. English maps rarely show the French names outside of Quebec, which itself is spelled *Québec* in French.

The study of placenames is called toponymy, while that of the origin and historical usage of placenames as words is etymology.

In order to improve legibility or to aid the illiterate, some maps have been produced using pictograms to represent places. The iconic example of this practice is Lance Wyman's early plans for the Mexico City Metro, on which stations were shown simply as stylized logos. Wyman also prototyped such a map for the Washington Metro, though ultimately the idea was rejected. Other cities experimenting with such maps are Fukuoka, Guadalajara and Monterrey.

Map symbology

Cartographic symbology encodes information on the map in ways intended to convey information to the map reader efficiently, taking into consideration the limited space on the map, models of human understanding through visual means, and the likely cultural background and education of the map reader. Symbology may be implicit, using universal elements of design, or may be more specific to cartography or even to the map.

A map may have any of many kinds of symbolization. Some examples are:

- A legend, or key, explains the map's pictorial language.
- A title indicates the region and perhaps the theme that the map portrays.
- A neatline frames the entire map image.
- A compass rose or north arrow provides orientation.
- An overview map gives global context for the primary map.
- A bar scale translates between map measurements and real distances.
- A map projection provides a way to represent the curved surface on the plane of the map.

The map may declare its sources, accuracy, publication date and authorship, and so forth. The map image itself portrays the region.

Map colouring is another form of symbology, one whose importance can reach beyond aesthetic. In complex thematic maps, for example, the colour scheme's structure can critically affect the reader's ability to understand the map's information. Modern computer displays and print technologies can reproduce much of the gamut that humans can perceive, allowing for intricate exploitation of human visual discrimination in order to convey detailed information.

Quantitative symbols give a visual indication of the magnitude of the phenomenon that the symbol represents. Two major classes of symbols are used to portray quantity. Proportional symbols change size according to phenomenon's magnitude, making them appropriate for representing statistics. Choropleth maps portray data collection areas, such as counties or census tracts, with colour. Using colour this way, the darkness and intensity (or value) of the colour is evaluated by the eye as a measure of intensity or concentration.

Map key or legend

The map key, or legend, describes how to interpret the map's symbols and may give details of publication and authorship.

Examples of point symbols

Map labeling

Most maps label features so that the map reader can know features' names. For example, country names may be printed on a world map, each label within the outline of the country it names. Features and background may be in any colour, which can make reading labels printed over them harder to read than reading text from a book. Two traits of good labels are legibility and easy feature association. In order for a label to be legible, it must have a type, size and colour that are easy to read. Ideally, a label would not interfere with other map features or labels. A halo may be placed around the label to contrast it against the background. A label must also be easily associated with the feature it names, regardless of the feature's category or extent. Choosing a location for a label to enhance this association while avoiding clutter and conflict with other labels is an art that incorporates many techniques. One technique is to use a different font per category of feature. For example, using an italicized, dark blue font for a water label can suggest waves. Or, labels for contour lines can be thin and have the same colour as the contours. In difficult cases where there is not enough space to place the label near the feature to form an unambiguous association, a leader line can connect the label to the feature.

Map generalization

A good map has to compromise between portraying the items of interest (or themes) in the right place on the map, and the need to show that item using text or a symbol, which take up space on the map and might displace some other item of information. The cartographer is thus constantly making judgements about what to include, what to leave out and what to show in a *slightly* incorrect place. This issue assumes more importance as the scale of the map gets smaller (i.e. the map shows a larger area) because the information shown on the map takes up more space *on the ground*. A good example from the late 1980s was the Ordnance Survey's first digital maps, where the *absolute* positions of major roads were sometimes a scale distance of hundreds of meters away from ground truth, when shown on digital maps at scales of 1:250,000 and 1:625,000, because of the overriding need to annotate the features.

Map projections

The Earth being spherical, any flat representation generates distortions such that shapes and areas cannot both be conserved simultaneously, and distances can never all be preserved. The mapmaker must choose a suitable *map projection* according to the space to be mapped and the purpose of the map.

Cartographic errors

Some maps contain deliberate errors or distortions, either as propaganda or as a "watermark" to help the copyright owner identify infringement if the error appears in competitors' maps. The latter often come in the form of nonexistent, misnamed, or misspelled "trap streets". Other names and forms for this are paper townsites, fictitious entries, and copyright easter eggs.

Another motive for deliberate errors is cartographic "vandalism": a mapmaker wishing to leave his or her mark on the work. Mount Richard, for example, was a fictitious peak on the Rocky Mountains' continental divide that appeared on a Boulder County, Colourado map in the early 1970s. It is believed to be the work of draftsman Richard Ciacci. The fiction was not discovered until two years later.

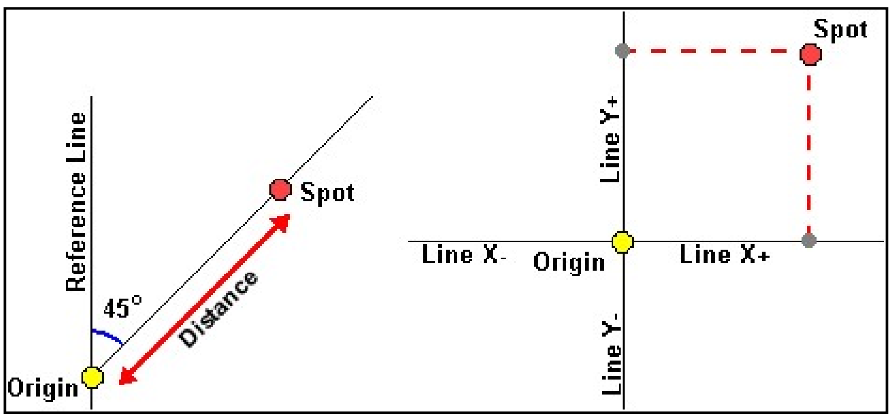
Sandy Island (New Caledonia) is an example of a fictitious location that stubbornly survives, reappearing on new maps copied from older maps while being deleted from other new editions.

UNIT 2: NATURE AND TYPES OF GEOID; CONCEPT OF DATUM WITH SPECIAL REFERENCE TO NAD, EVEREST AND WGS-84; PRINCIPLES OF SPHERICAL TRIGONOMETRY

Imagine yourself standing in a totally flat, featureless expanse where everywhere is the same as everywhere else. If someone asked, "Where are you?" what would that mean? Where in relation to what? Location assumes a reference point; in fact, you are lost without one.

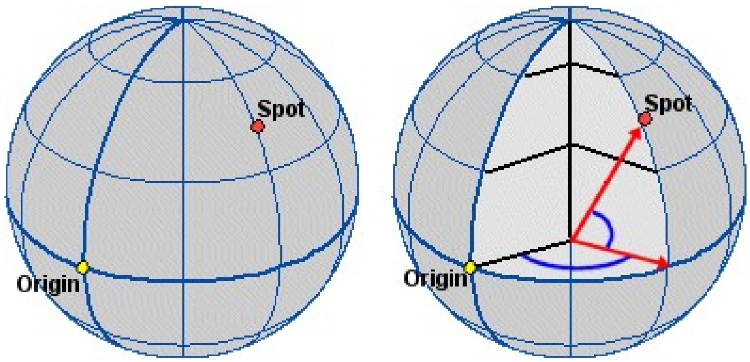
This is the problem—lack of orientation in space—that coordinate systems solve. They establish the whereabouts of places by defining a starting point and a way to measure distances and directions from that point.

The starting point is called the origin of the system. On a flat surface, if you have an origin, a distance measure of some kind (footsteps, for example), and a way to fix your direction, you can describe locations uniquely.



Two ways to identify locations on a plane surface: Left: Draw a straight line (reference line) through the origin. Draw another straight line from the origin to a point (spot). The angle between the two lines and the distance from origin to spot identifies the point uniquely. Right: Draw two straight lines through the origin that intersect each other at a 90-degree angle. Label the lines X and Y. Sub-label them + and - on opposite sides of the origin. A point's location can be uniquely identified by a distance from the origin parallel to X (+ or -) and a distance from the origin parallel to Y (+ or -).

What if the surface is not flat, but a sphere? The coordinate system is still based on an origin and two lines intersecting at right angles, but points are not identified in terms of surface distance. Instead, angles to the center of the sphere are used.



Left: The origin and a location to be identified. Right: The X and Y lines still intersect at the origin, but points are identified by interior angles of the sphere.

Geographic coordinate systems reference locations on the earth's surface using angular measurements, while projected coordinate systems reference locations on the earth's surface using linear measurements. In the latter method, the curved surface of the earth must be "projected" so that it can be represented as a flat surface. This module will focus on ways to determine appropriate coordinate systems and map projections for mapping problems.

Geographic Coordinate Systems

A geographic coordinate system (GCS) is a reference system that uses a three-dimensional spherical surface to identify points or areas on the surface of the earth. Each geographic coordinate system has three components:

- an angular unit of measure
- a prime meridian
- a datum, which includes a spheroid

The angular unit of measure is usually degrees, in which the unit represents one part in 360 of a circle. But it may also be grads, in which the unit represents one part in 400 of a circle.

The prime meridian defines the zero value for longitude. (It is unnecessary to define a zero value for latitude, because this is always the equator.)

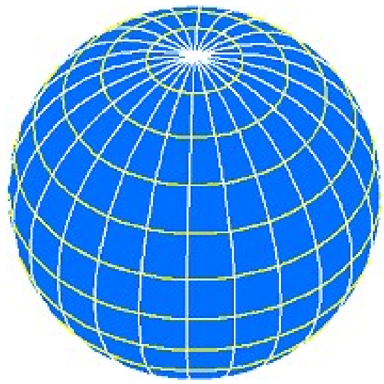
The datum specifies the position of the sphere or spheroid relative to the center of the earth. It defines the origin and orientation of latitude and longitude lines.

Any spatial data you add to ArcMap is associated with a particular geographic coordinate system (GCS), distinguished by its unique set of components—especially by its datum.

Angular Units of Measure

The Latitude-Longitude System

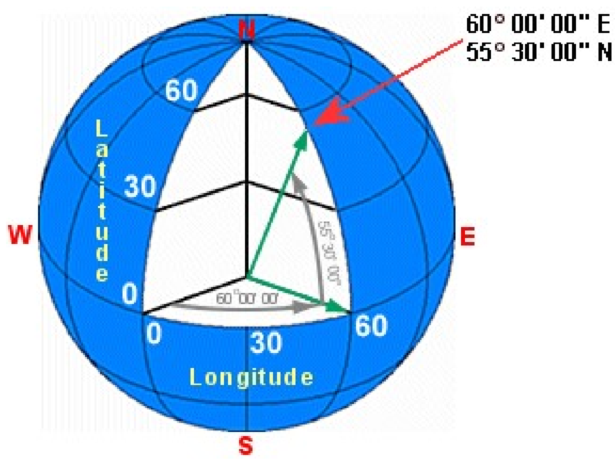
Lines of latitude are parallel circles that run east–west around the globe, getting smaller as they approach the poles. Their values range from +90° (North Pole) to -90° (South Pole), with the equator being 0°. Lines of longitude are half-circles that run north–south from pole to pole. Their values range from 0° to 360°, with 0° being arbitrarily, but commonly designated as the line that passes through Greenwich, England.



The network of intersecting lines is called a graticule.

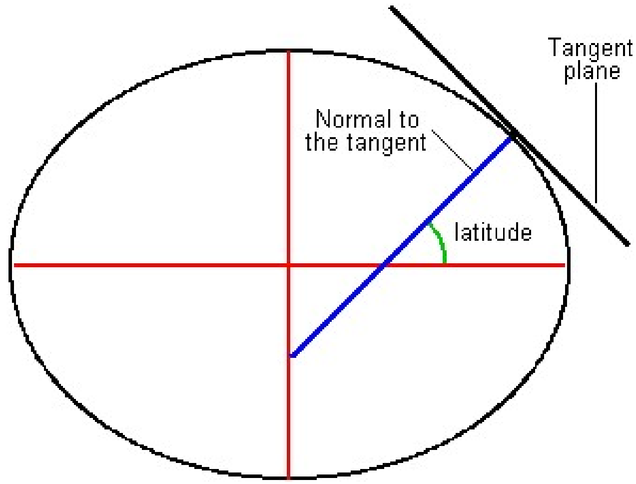
Defining Latitude and Longitude

When the earth is treated as a sphere, latitude is the angle between two lines drawn to the center of the earth: one from a point on the equator and the other from a specified point on the same meridian. Longitude is an angle between two lines drawn to the center of the earth from the equator: one at the prime meridian and the other at a specified meridian.



Latitude and longitude defined on a sphere.

When the earth is treated as a spheroid, latitude is the angle between two lines drawn toward the center of the earth: one from a point on the equator and the other normal (perpendicular) from a plane that touches the spheroid at a specified point.



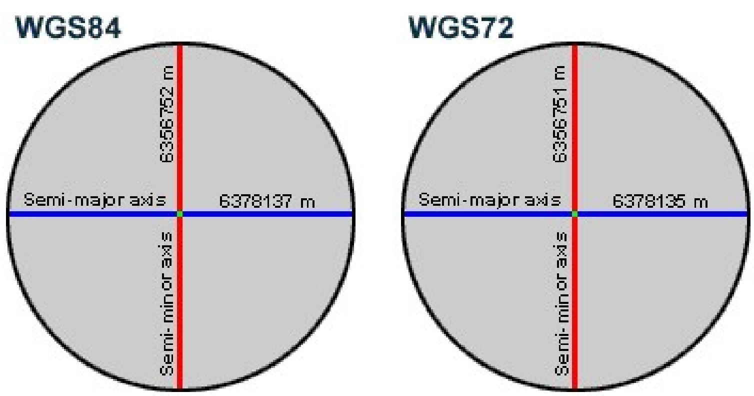
Latitude defined on a spheroid. The lines do not intersect at the center of the earth (unless the plane is tangent at a pole).

Note: Longitude is defined on a spheroid the same as it is on a sphere.

Spheroids

The surface of the earth's gravity field is called the geoid. The shape of the geoid is irregular, but overall, it is approximately the same as mean sea level. Various spheroids have been devised to simplify the model of the geoid and make it more uniform. Each geographic coordinate system includes a spheroid.

A particular spheroid can be selected for use in a specific geographic area because that particular spheroid does an exceptionally good job of mimicking the geoid for that part of the world.



A spheroid is simply an ellipsoid that approximates a sphere. These examples are two common world spheroids in use today with their values rounded to the nearest meter. For each spheroid, the difference between its major axis and its minor axis is less than 0.34 percent.

A datum is built on top of the selected spheroid, and it can incorporate local variations in elevation. Spheroids create a totally smooth surface across the world, but because this does not reflect reality very well, this ability of a local datum to incorporate local variations in elevation is important.

Common Spheres and Spheroids

The most commonly used radius for the earth is 6,371 kilometers. This is the default sphere used by ArcGIS for sphere-based map projections.

Of the twenty or more spheroids defined for the earth since 1800, several are still in use today. The WGS84 spheroid, determined by satellite, is becoming a global standard, but a huge amount of spatial data—digital and nondigital—is based on other spheroids. Here are the dimensions for a few widely used spheroids. You can see that the differences among them are not that significant.

Name	Date	Semi-major axis (meters)	Semi-minor axis (meters)	Flattening Ratio
WGS84	1984	6,378,137	6,356,752.3	1/298.257
GRS80	1980	same	same	same
WGS72	1972	6,378,135	6,356,750.5	1/298.26
International	1924	6,378,388	6,356,911.9	1/297
Clarke 1866	1866	6,378,206.4	6,356,583.8	1/294.98

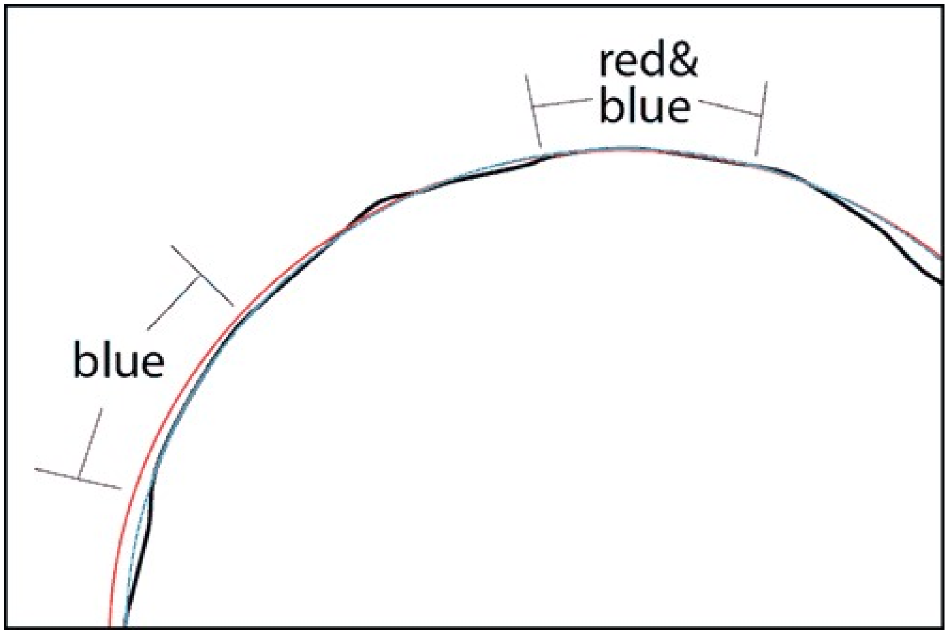
Note: The terms "spheroid" and "ellipsoid" are often used interchangeably and have caused confusion for many GIS users. Most coordinate system authorities consider both terms equally correct. Because spheroid is a bit more descriptive than ellipsoid, and because Esri software and documentation use spheroid, spheroid is the term that will be used in this course.

Datums

Every map projection and coordinate system begins with a precisely surveyed starting point. The starting point and the network of points that extends from it is called the datum. A datum provides a frame of reference for measuring locations on the surface of the earth. It defines the origin and orientation of latitude and longitude lines.

Each datum has a name and often a date associated with it, such as European Datum of 1950, the Pulkovo Datum of 1942, or the world-wide datum WGS 1984. Every datum is based on a specific spheroid because surveyors selected the spheroid that best represented their part of the earth when they defined a new datum. This means that every datum has an associated spheroid. While a spheroid approximates the shape of the earth, a datum defines the position of the spheroid relative to the center of the earth. The underlying datum and spheroid to which coordinates for a dataset are referenced can change the coordinate values.

As you probably can guess from some of the names above (European, Pulkovo, WGS), most datums were originally designed to be used only in a specific country or region of the globe. Using a local datum outside its designated area will lead to serious errors.



Although highly exaggerated, this graphic illustrates that the earth itself (the black line) is irregularly shaped. The blue spheroid works well in two areas, but not over the entire surface of the earth. The red spheroid works well in only one area, but it may be a better fit there than the blue spheroid. With the advent of global positioning systems (GPS), new datums and ellipsoids have been developed for the entire globe. WGS 84 is now accepted as a universal global datum.

The elements of a datum are:

- a spheroid
- an initial reference point, or origin
- an azimuth from the origin to a second point
- the orientation of the spheroid to the geoid (This is defined as the distance separating the geoid and the spheroid at the origin, and it is usually zero.)

Datum Transformations

A new data frame has no coordinate system. When you add your first layer to it, the data frame adopts that layer's coordinate system. For any subsequent layers that you add, one of two things happens.

- If the new layer's coordinate system already matches the data frame's, the layer is added without question.
- If the new layer's coordinate system is different, you get a geographic coordinate system warning and ArcMap changes the coordinate system of the new layer to match the first one. This operation is called a datum transformation (or coordinate system transformation).

For all new layers that you add, ArcMap performs these transformations automatically so that layers with different geographic coordinate systems can be displayed together. The transformations are done on the fly, meaning they are applied only inside the data frame; the geographic coordinate systems of the datasets on disk are not changed.

Although ArcMap will do datum transformations for you, complications may arise when there are multiple versions of a transformation. For example, when there are several formulas that convert from one datum to another, each may work best for a particular region. For each datum transformation, ArcMap has to choose a default formula. If your data happens to lie in a region other than the one for which the chosen transformation was designed, your features may not line up properly until you override the default transformation and assign the optimal one.

UNIT – 3: PRINCIPLES AND PROPERTIES OF UTM PROJECTIONS

The **Universal Transverse Mercator** (**UTM**) is a system for assigning coordinates to locations on the surface of the Earth. Like the traditional method of latitude and longitude, it is a horizontal position representation, which means it ignores altitude and treats the earth as a perfect ellipsoid. However, it differs from global latitude/longitude in that it divides earth into 60 zones and projects each to the plane as a basis for its coordinates. Specifying a location means specifying the zone and the x, y coordinate in that plane. The projection from spheroid to a UTM zone is some parameterization of the transverse Mercator projection. The parameters vary by nation or region or mapping system.

Most zones in UTM span 6 degrees of longitude, and each has a designated central meridian. The scale factor at the central meridian is specified to be 0.9996 of true scale for most UTM systems in use.

History

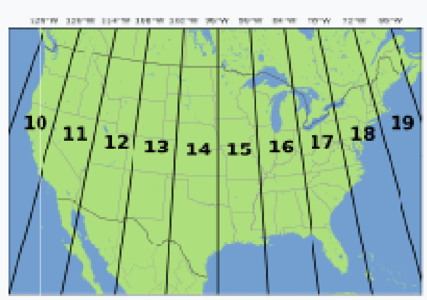
The National Oceanic and Atmospheric Administration (NOAA) website states the system to have been developed by the United States Army Corps of Engineers, starting in the early 1940s. However, a series of aerial photos found in the Bundesarchiv-Militärarchiv (the military section of the German Federal Archives) apparently dating from 1943–1944 bear the inscription UTMREF followed by grid letters and digits, and projected according to the transverse Mercator, a finding that would indicate that something called the UTM Reference system was developed in the 1942–43 time frame by the Wehrmacht. It was probably carried out by the AbteilungfürLuftbildwesen (Department for Aerial Photography). From 1947 onward the US Army employed a very similar system, but with the now-standard 0.9996 scale factor at the central meridian as opposed to the German 1.0. For areas within the contiguous United States the Clarke Ellipsoid of 1866 was used. For the remaining areas of Earth, including Hawaii, the International Ellipsoid was used. The World Geodetic System WGS84 ellipsoid is now generally used to model the Earth in the UTM coordinate system, which means current UTM northing at a given point can differ up to 200 meters from the old. For different geographic regions, other datum systems can be used.

Prior to the development of the Universal Transverse Mercator coordinate system, several European nations demonstrated the utility of grid-based conformal maps by mapping their territory during the interwar period. Calculating the distance between two points on these maps could be performed more easily in the field (using the Pythagorean theorem) than was possible using the trigonometric formulas required under the graticule-based system of latitude and longitude. In the post-war years, these concepts were extended into the Universal Transverse Mercator/Universal Polar Stereographic (UTM/UPS) coordinate system, which is a global (or universal) system of grid-based maps.

The transverse Mercator projection is a variant of the Mercator projection, which was originally developed by the Flemish geographer and cartographer Gerardus Mercator, in 1570. This projection is conformal, which means it preserves angles and therefore shapes across small regions. However, it distorts distance and area.

Definition

UTM zone



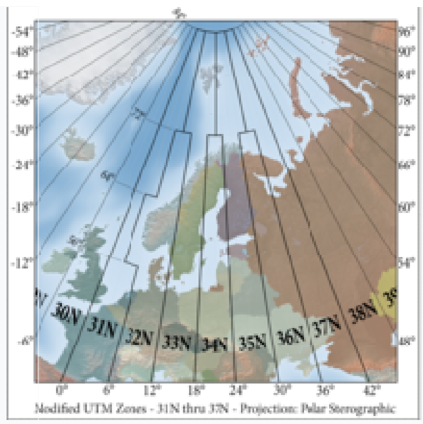
Simplified view of contiguous US UTM zones, projected with Lambert conformal conic.

The UTM system divides the Earth into 60 zones, each 6° of longitude in width. Zone 1 covers longitude 180° to 174° W; zone numbering increases eastward to zone 60, which covers longitude 174°E to 180°. The polar regions south of 80°S and north of 84°N are excluded.

Each of the 60 zones uses a transverse Mercator projection that can map a region of large north-south extent with low distortion. By using narrow zones of 6° of longitude (up to 668 km) in width, and reducing the scale factor along the central meridian to 0.9996 (a reduction of 1:2500), the amount of distortion is held below 1 part in 1,000 inside each zone. Distortion of scale increases to 1.0010 at the zone boundaries along the equator.

In each zone the scale factor of the central meridian reduces the diameter of the transverse cylinder to produce a secant projection with two standard lines, or lines of true scale, about 180 km on each side of, and about parallel to, the central meridian (Arc cos 0.9996 = 1.62° at the Equator). The scale is less than 1 inside the standard lines and greater than 1 outside them, but the overall distortion is minimized.

Overlapping grids



Universal Transverse Mercator (UTM) Grid Zones 31N thru 37N differ from the standard 6° wide by 84° zone for the northern hemisphere, in part to accommodate the southern half of the Kingdom of Norway. For more on its history, see Clifford J. Mugnier's article on Grids &Datums of The Kingdom of Norway that appeared in the October 1999 issue of PE&RS http://www.asprs.org/a/resources/grids/10-99-norway.pdf

Distortion of scale increases in each UTM zone as the boundaries between the UTM zones are approached. However, it is often convenient or necessary to measure a series of locations on a single grid when some are located in two adjacent zones. Around the boundaries of large scale maps (1:100,000 or larger) coordinates for both adjoining UTM zones are usually printed within a minimum distance of 40 km on either side of a zone boundary. Ideally, the coordinates of each position should be measured on the grid for the zone in which they are located, but because the scale factor is still relatively small near zone boundaries, it is possible to overlap measurements into an adjoining zone for some distance when necessary

Latitude bands

Latitude bands are not a part of UTM, but rather a part of the military grid reference system (MGRS). They are however sometimes used.

Latitude bands

Each zone is segmented into 20 latitude bands. Each latitude band is 8 degrees high, and is lettered starting from "C" at 80°S, increasing up the English alphabet until "X", omitting the letters "I" and "O" (because of their similarity to the numerals one and zero). The last latitude band, "X", is extended an extra 4 degrees, so it ends at 84°N latitude, thus covering the northernmost land on Earth.

Latitude bands "A" and "B" do exist, as do bands "Y" and "Z". They cover the western and eastern sides of the Antarctic and Arctic regions respectively. A convenient mnemonic to remember is that the letter "N" is the first letter in "northern hemisphere", so any letter coming before "N" in the alphabet is in the southern hemisphere, and any letter "N" or after is in the northern hemisphere.

Notation

The combination of a zone and a latitude band defines a grid zone. The zone is always written first, followed by the latitude band. For example, (see image, top right), a position in Toronto, Ontario, Canada, would find itself in zone 17 and latitude band "T", thus the full grid zone reference is "17T". The grid zones serve to delineate irregular UTM zone boundaries. They also are an integral part of the military grid reference system.

A note of caution: A method also is used that simply adds N or S following the zone number to indicate North or South hemisphere (the easting and northing coordinates along with the zone number supplying everything necessary to geolocate a position except which hemisphere). However, this method has caused some confusion since, for instance, "50S" can mean southern hemisphere but also *grid zone* "50S" in the northern hemisphere. There are many possible ways to disambiguate between the two methods, two of which are demonstrated later in this article

Exceptions

These grid zones are uniform over the globe, except in two areas. On the southwest coast of Norway, grid zone 32V (9° of longitude in width) is extended further west, and grid zone 31V (3° of longitude in width) is correspondingly shrunk to cover only open water. Also, in the region around Svalbard, the four grid zones 31X (9° of longitude in width), 33X (12° of longitude in width), 35X (12° of longitude in width), and 37X (9° of longitude in width) are extended to cover what would otherwise have been covered by the seven grid zones 31X to 37X. The three grid zones 32X, 34X and 36X are not used.

Locating a position using coordinates

A position on the Earth is given by the UTM zone number and the easting and northing planar coordinate pair in that zone. The point of origin of each UTM zone is the intersection of the equator and the zone's central meridian. To avoid dealing with negative numbers, the central meridian of each zone is defined to coincide with 500000 meters East. In any zone a point that has an easting of 400000 meters is about 100 km west of the central meridian. For most such points, the true distance would be slightly more than 100 km as measured on the surface of the Earth because of the distortion of the projection. UTM eastings range from about 167000 meters to 833000 meters at the equator.

In the northern hemisphere positions are measured northward from zero at the equator. The maximum "northing" value is about 9300000 meters at latitude 84 degrees North, the north end of the UTM zones. In the southern hemisphere northings decrease southward from the equator to about 1100000 meters at 80 degrees South, the south end of the UTM zones. The northing at the equator is set at 10000000 meters so no point has a negative northing value.

The CN Tower is at 43°38′33.24″N 79°23′13.7″W, which is in UTM zone 17, and the grid position is 630084 m east, 4833438 m north. Two points in Zone 17 have these coordinates, one in the northern hemisphere and one in the south; one of two conventions is used to say which:

1. Append a hemisphere designator to the zone number, "N" or "S", thus "17N 630084 4833438". This supplies the minimum information to define the position uniquely.
2. Supply the grid zone, i.e., the latitude band designator appended to the zone number, thus "17T 630084 4833438". The provision of the latitude band along with northing supplies redundant information (which may, as a consequence, be contradictory if misused).

Because latitude band "S" is in the northern hemisphere, a designation such as "38S" is unclear. The "S" might refer to the latitude band (32°N–40°N) or it might mean "South". It is therefore important to specify which convention is being used, e.g., by spelling out the hemisphere, "North" or "South", or using different symbols, such as − for south and + for north.

UNIT 4: CONCEPT OF GEOINFORMATICS; REMOTE SENSING PLATFORMS AND SENSORS

Concept of Geoinformatics

Having introduced the 5D datum paradigm that needs to be adequately dealt with to define geodata accurately, consistently, timely and completely so that it can be used without any restrictions in space, time or scale and further, having appreciated the truly global dimension of the digital Earth, to put everything in perspective, it is now appropriate to focus on geoinformatics. Like for all other disciplines elaborated in this book it is only right to begin this discussion with pertinent definitions.

Although geoinformatics is a fairly recent terminology, various definitions of the same have been advanced by different authors. For instance, Raju (2003) describes geoinformatics as "the science and technology dealing with the structure and character of spatial information, its capture, its classification and qualification, its storage, processing, portrayal and dissemination, including the infrastructure necessary to secure optimal use of this information". Similarly, Ehlers (2003) defines geoinformatics as "the art, science or technology dealing with the acquisition, storage, processing, production, presentation and dissemination of geoinformation".

The bottom line is that there is no globally accepted definition of geoinformatics. However, as a multidisciplinary field, geoinformatics has at its core different technologies that support the acquisition, analysis and visualization of geodata. The geodata is usually acquired from Earth observation sensors as remotely sensed images, analyzed by geographic information systems (GIS) and visualized on paper or on computer screens. Furthermore, it combines geospatial analysis and modeling, development of geospatial databases, information systems design, human-computer interaction and both wired and wireless networking technologies. Geoinformatics uses geocomputation and geovisualization for analyzinggeoinformation. Typical branches of geoinformatics include: *cartography, geodesy, geographic informationsystems, global navigation satellite systems (GNSS), photogrammetry, remote sensing,and web mapping*. These different disciplines that have been developed over different time epochs form the main subject matter of this book.

By combining the ever-increasing computational power, modern telecommunications technologies, abundant and diverse geodata, and more advanced image analysis algorithms available, and integrating technologies such as remote sensing, GIS and GNSS, many opportunities for application of geoinformatics have been realized. Today, many applications routinely benefit from geoinformatics including; urban planning and land use management, in-car navigation systems, virtual globes, public health, local and national gazetteer management, environmental modelling and analysis,military, transport network planning and management, agriculture, meteorology and climate change, oceanography and coupled ocean and atmosphere modeling, business location planning, architecture and archaeological reconstruction, telecommunications, criminology and crime simulation, aviation and maritime transport etc.

Consequently, geoinformatics has become a very important technology to decision-makers across a wide range of disciplines, industries, commercial sector, environmental agencies, local and national government, research and academia, national survey and mapping organizations, international organizations, United Nations, emergency services, public health and epidemiology, crime mapping, transportation and infrastructure, information technology industries, GIS consulting firms, environmental management agencies, tourist industry, utility companies, market analysis and e-commerce, mineral exploration etc. Increasingly, many government and non government agencies worldwide are using geodata and geoinformatics for managing their day to day activities. Figure 2.2 shows a conceptual framework that underlines the role of geoinformatics in supporting spatial decision-making (Awange and Kiema 2013).

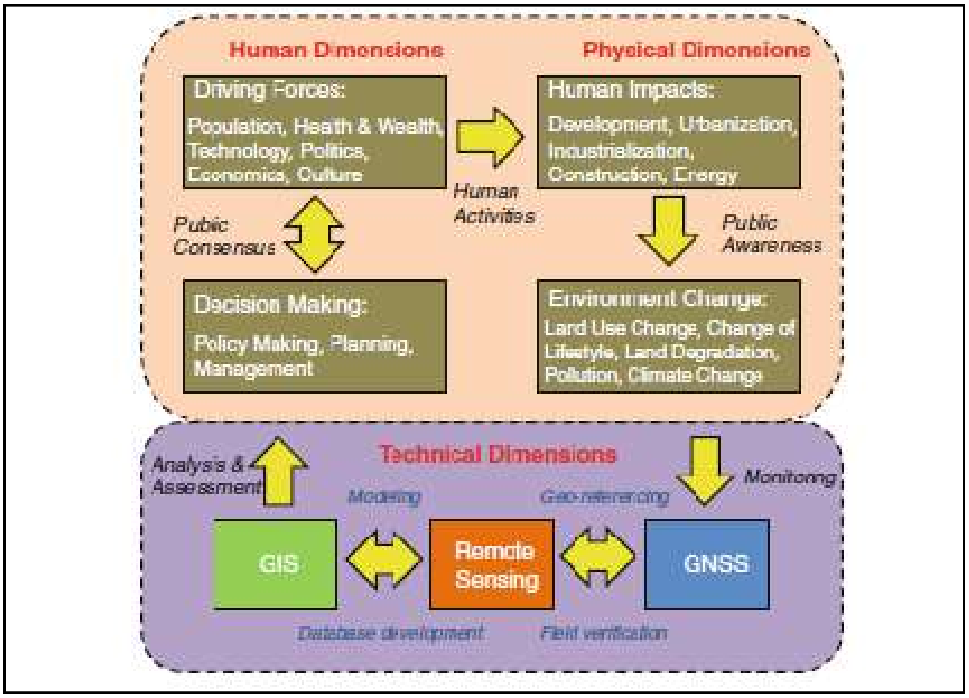


Figure: 4.1 Conceptual framework showing the role of geoinformatics in spatial decision support (Modified after Murari 1999)

REMOTE SENSING PLATFORMS AND SENSORS

Remote Sensing Platforms

Imaging in remote sensing can be carried out from both satellite and aircraft platforms. In many ways their sensors have similar characteristics although differences in their altitude and stability can lead to very different image properties.

There are essentially two broad classes of satellite program: those satellites that sit at geostationary altitudes above the earth's surface and which are generally associated with weather and climate studies, and those which orbit much closer to the earth's surface and that are generally used for earth surface and oceanographic observations. Usually, the low earth orbiting satellites are in a sun-synchronous orbit, in that their orbital plane precesses around the earth at the same rate that the sun appears to move across the earth's surface. In this manner the satellite acquires data at about the same local time on each orbit.

Low earth orbiting satellites can also be used for meteorological studies. Notwithstanding the differences in altitude, the wavebands used for the geostationary and the low earth orbiting satellites, and for weather and earth observation satellites, are very comparable. The major distinction in the image data they provide generally lies in the spatial resolutions available. Whereas data acquired for earth resources purposes generally has pixel sizes of less than 100 m, that used for meteorological purposes (both at geostationary and lower altitudes) has a much coarser pixel, often of the order of 1 km.

The imaging technologies utilised in satellite programs have ranged from traditional cameras to mechanical scanners that record images of the earth's surface by moving the instantaneous field of view of the instrument across the earth's surface to record the upwelling energy.

Some weather satellites scan the earth's surface using the spin of the satellite itself while the sensor's pointing direction is varied (at a slower rate) along the axis of the satellite. The image data is then recorded in a raster-scan fashion not unlike that used for the production of television pictures.

A more common image recording mechanism, used in the Landsat program, has been to carry a mechanical scanner that records at right angles to the direction of the satellite motion to produce raster-scans of data. The forward motion of the vehicle then allows an image strip to be built up from the raster-scans. That process is depicted in Fig. 1.6.

More recent technology utilises a “push-broom" mechanism in which a linear imaging array with sufficient detectors is carried on the satellite, normal to the satellite’s motion, such that each pixel can be recorded individually. The forward motion of the satellite then allows subsequent pixels to be recorded along the satellite travel direction in the manner shown in Fig. 1.7. As might be expected, the time over which the energy emanating from the earth’s surface per pixel is larger with push broom scanning than for the mechanical scanners, generally allowing finer spatial resolutions to be achieved.

Aircraft scanners operate with essentially the same principles as those found on satellites. Both mechanical scanners and CCD arrays are commonly employed.

An interesting development in the past decade has been to employ rectangular detector arrays which, in principle, could be used to capture a two dimensional image underneath the satellite. They are normally used, however, to record pixels in the across track direction, as with push broom scanners, with the other dimension employed to record many spectral channels of data simultaneously. This is depicted in Fig. 1.8. Often as many as 200 or so channels are recorded in this manner so that a very good rendition of the spectra depicted in Fig. 1.3 can be obtained. As a result the devices are often referred to as imaging spectrometers and the data described as *hyperspectral*, as against multispectral when of the order of 10 wavebands are recorded. Figure 1.9 shows the quality of the spectral data per pixel possible with an imaging spectrometer, compared with the detail obtainable from the Landsat MSS and TM instruments.

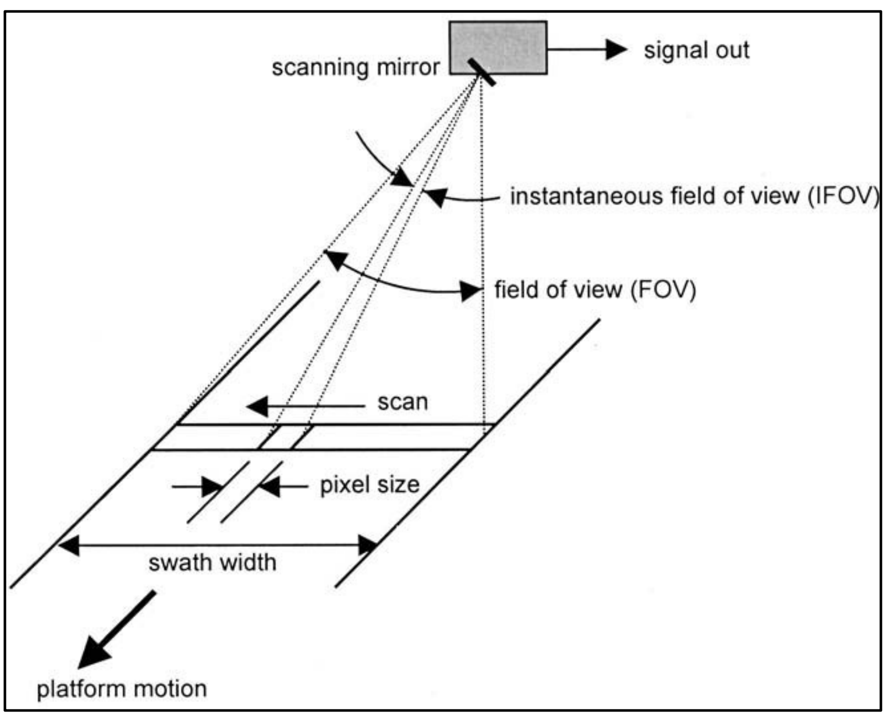


Fig. 1.6. Image formation by mechanical line scanning (Richards and Xiuping 2006)

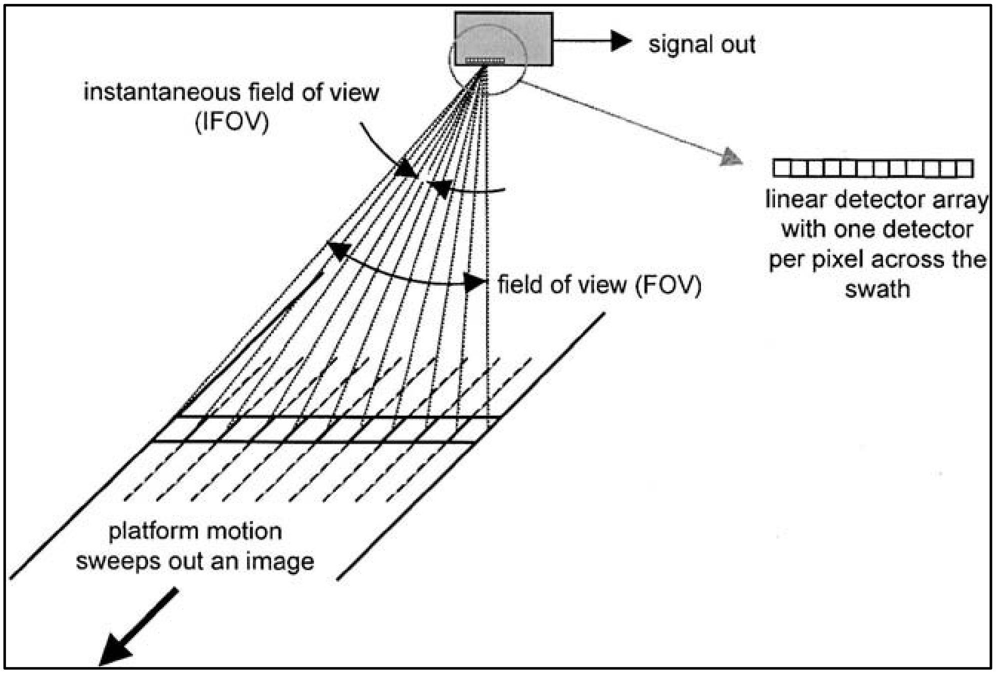


Fig. 1.7. Push broom line scanning in the along-track direction (Richards and Xiuping 2006)

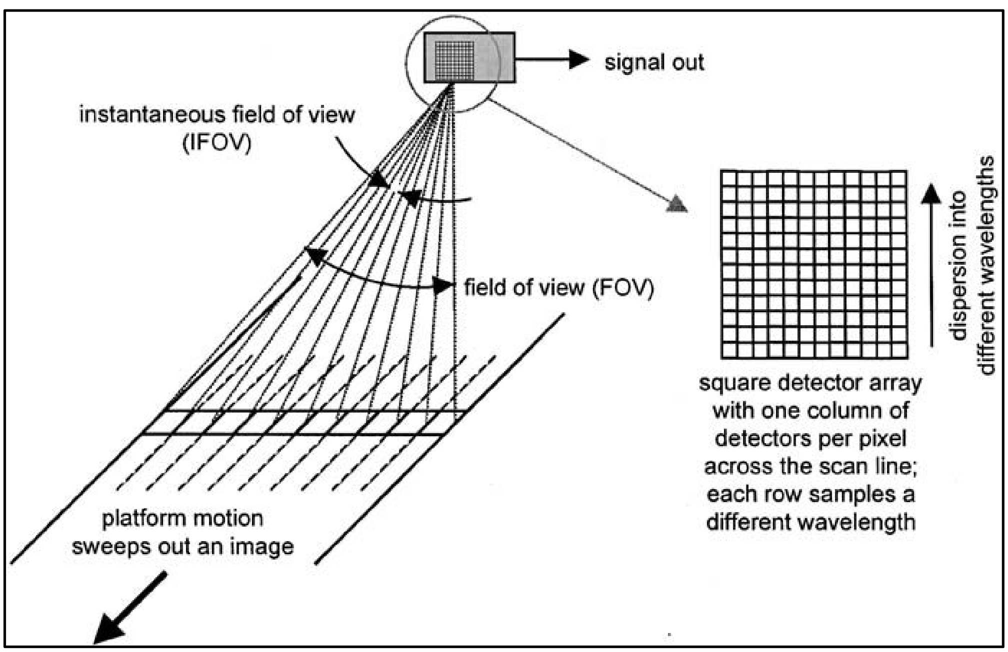


Figure 1.8 Use of square detector array to achieve along track line scanning and the recording of many spectral measurements simultaneously (Richards and Xiuping 2006)

IMAGE DATA SOURCES IN THE MICROWAVE REGION

Side Looking Airborne Radar and Synthetic Aperture Radar

Remote sensing image data in the microwave range of wavelengths is generally gathered using the technique of side-looking radar, as illustrated in Fig. 1.10. When used with aircraft platforms it is more commonly called SLAR (side looking airborne radar), a technique that requires some modification when used from spacecraft altitudes, as discussed in the following.

In SLAR a pulse of electrical energy at the microwave frequency (or wavelength) of interest is radiated to the side of the aircraft at an incidence angle of θi. By the same principle as radars used for air navigation and shipping, some of this transmitted energy is scattered from the ground and returned to the receiver on the aircraft. The time delay between transmission and reflection identifies the slant distance to the “target" from the aircraft, while the strength of the return contains information on the so-called scattering coefficient of the target region of the earth’s surface. The actual received signal from a single transmitted pulse consists of a continuum of reflections from the complete region of ground actually illuminated by the radar antenna. In Fig. 1.10 this can be identified as the range beamwidth of the antenna. This is chosen at design to give a relation between swath width and altitude, and tends to be rather broad. By comparison the along-track, or so-called azimuth, beamwidth is chosen as small as possible so that the reflections from a single transmitted pulse can be regarded as having come from a narrow strip of terrain broadside to the aircraft. The forward velocity of the aircraft is then arranged so that the next transmitted pulse illuminates the next strip of terrain along the swath. In this manner the azimuth beamwidth of the antenna defines the spatial resolution in the azimuth direction whereas the time resolution possible between echos from two adjacent targets in the range direction defines the spatial resolution in the slant direction.

From an image product viewpoint the slant range resolution is not of interest. Rather it is the projection of this onto the horizontal plane as ground range resolution that is of value to the user. A little thought reveals that the ground range resolution is better at larger incidence angles and thus on the far side of the swath; it can be shown that the ground range size of a resolution element (pixel) is given by

$$r_g = c\tau/2 \sin \theta_i$$

whereτ is the length of the transmitted pulse and c is the velocity of light. (Often a simple pulse is not used. Instead a so-called linear chirped waveform is transmitted and signal processing on reception is used to compress this into a narrow pulse. For the present discussion however it is sufficient to consider the transmitted waveform to be a simple pulse or burst of the frequency of interest.)

The azimuth size of a resolution element is related to the length (or aperture) of the transmitting antenna in the azimuth direction, l, the wavelength λ and the range R_0 between the aircraft and the target, and is given by

$$r_a = R_0\lambda / l$$

This expression shows that a 10 m antenna will yield an azimuth resolution of 20 m at a slant range of 1 km for radiation with a wavelength of 20 cm. However if the slant range is increased to say 100 km – i.e. at low spacecraft altitudes – then a 20 m azimuth resolution would require an antenna of 1 km length, which clearly is impracticable.

Therefore when radar image data is to be acquired from spacecraft, a modification of SLAR referred to as synthetic aperture radar (SAR) is used. Essentially this utilizes the motion of the space vehicle, during transmission of the ranging pulses, to give an effectively long antenna, or a so-called synthetic aperture. This principle is illustrated in Fig. 1.11, wherein it is seen that an intentionally large azimuth beamwidth is employed to ensure that a particular spot on the ground is illuminated and thus provides reflections over a length of spacecraft travel equivalent to the synthetic aperture required.

A discussion of the details of the synthetic aperture concept and the signal processing required to produce a high azimuth resolution is beyond the scope of this treatment. The matter is pursued further in Ulaby, Moore and Fung (1982), Elachi et al. (1982), Tomiyasu (1978), and Elachi (1983, 1988) (Richards and Xiuping 2006).

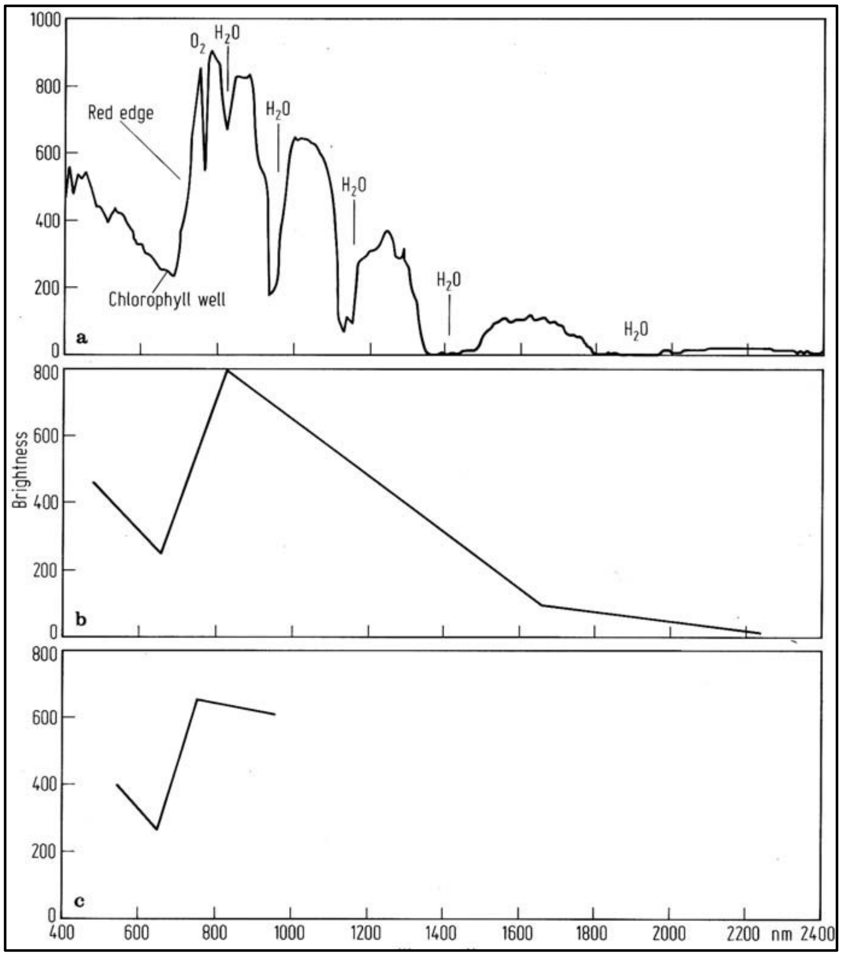


Fig. 1.9. Vegetation spectrum recorded by AVIRIS at 10 nm spectral sampling **a**, along with equivalent TM **b** and MSS **c** spectra. In **a** the fine absorption features resulting from at- mospheric constituents are shown, along with features normally associated with vegetation spectra (Richards and Xiuping 2006)

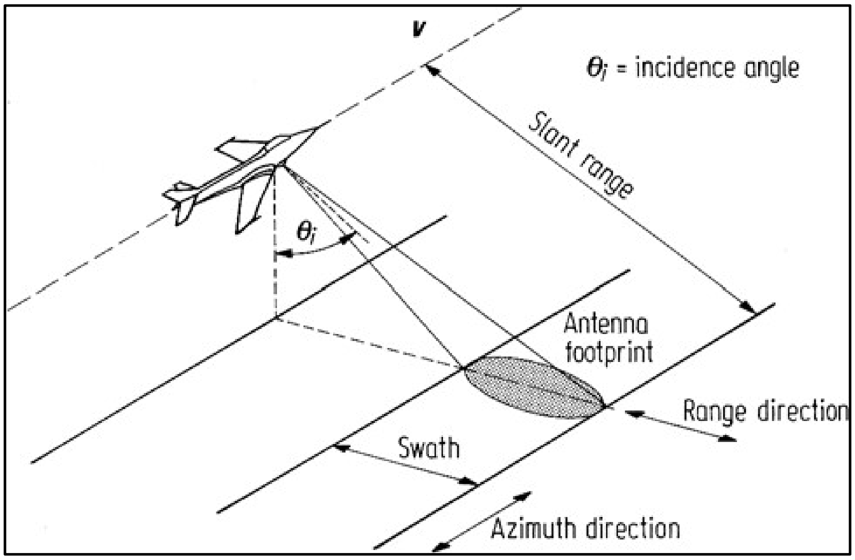


Fig. 1.10. Principle of side looking radar (Richards and Xiuping 2006)

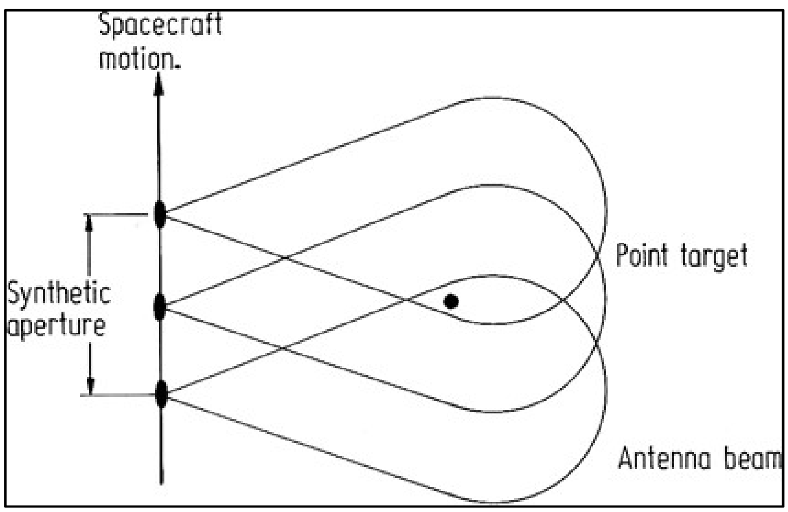


Fig. 1.11. The concept of synthesizing a large antenna by utilizing spacecraft motion along its orbital path. Here a view from above is shown, illustrating that a small real antenna is used to ensure a large real beamwidth in azimuth.As a consequence a point on the ground is illuminated by the full syntheticaperture (Richards and Xiuping 2006)

UNIT 5: NATURE OF EMR, EMS, AND INTERACTION WITH ATMOSPHERE AND SURFACE MATERIALS

ENERGY SOURCES AND RADIATION PRINCIPLES

Visible light is only one of many forms of electromagnetic energy. Radio waves, ultraviolet rays, radiant heat, and X-rays are other familiar forms. All this energy is inherently similar and propagates in accordance with basic wave theory. As shown in Figure 1.2, this theory describes electromagnetic energy as traveling in a harmonic, sinusoidal fashion at the "velocity of light" c. The distance from one wave peak to the next is the wavelength λ, and the number of peaks passing a fixed point in space per unit time is the wave frequency υ.

From basic physics, waves obey the general equation (1.1)

$$c = \upsilon\lambda$$

Because c is essentially a constant (3×10^8 m/sec), frequency $\nu$and wavelength λ for any given wave are related inversely, and either term can be used to

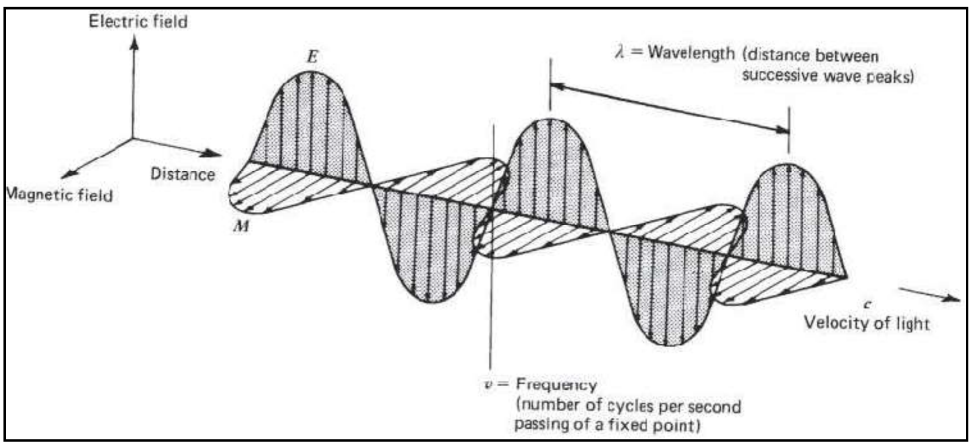


Figure 1.2 Electromagnetic wave. Components include a sinusoidal electric wave (E) and a similar magnetic wave (M) at right angles, both being perpendicular to the direction of propagation

characterize a wave. In remote sensing, it is most common to categorize electromagnetic waves by their wavelength location within the electromagnetic spectrum (Figure 1.3). The most prevalent unit used to measure wavelength along the spectrum is the micrometer(μm). A micrometer equals 1×10^{-6} m.

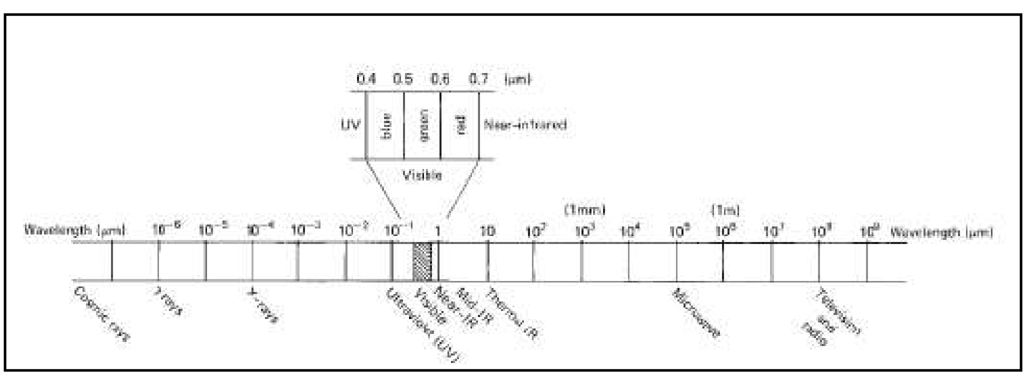


Figure 1.3 Electromagnetic spectrum

Although names (such as “ultraviolet” and “microwave”) are generally assigned to regions of the electromagnetic spectrum for convenience, there is no clear-cut dividing line between one nominal spectral region and the next. Divisions of the spectrum have grown from the various methods for sensing each type of radiation more so than from inherent differences in the energy characteristics of various wavelengths. Also, it should be noted that the portions of the electromagnetic spectrum used in remote sensing lie along a continuum characterized by magnitude changes of many powers of 10. Hence, the use of logarithmic plots to depict the electromagnetic spectrum is quite common. The “visible” portion of such a plot is an extremely small one, because the spectral sensitivity of the human eye extends only from about 0.4 μm to approximately 0.7 μm. The colour “blue” is ascribed to the approximate range of 0.4 to 0.5 μm, “green” to 0.5 to 0.6 μm, and “red” to 0.6 to 0.7μm. Ultraviolet (UV) energy adjoins the blue end of the visible portion of the spectrum. Beyond the red end of the visible region are three different categories of infrared (IR) waves: near IR (from 0.7 to 1.3 μm), midIR (from 1.3 to 3 μm; also referred to as shortwave IR or SWIR), and thermal IR (beyond 3 to 14 μm, sometimes referred to as longwave IR). At much longer wavelengths (1 mm to 1 m) is the microwave portion of the spectrum.

Most common sensing systems operate in one or several of the visible, IR, or microwave portions of the spectrum. Within the IR portion of the spectrum, itshould be noted that only thermal-IR energy is directly related to the sensation ofheat; near- and mid-IR energy are not.

Although many characteristics of electromagnetic radiation are most easily described by wave theory, another theory offers useful insights into how electromagnetic energy interacts with matter. This theory—the particle theory—suggests that electromagnetic radiation is composed of many discrete units called photons or quanta. The energy of a quantum is given as (1.2)

$$Q = hv$$

where

Q = energy of a quantum*;* joules (J)

h = Planck’s constant, 6.626×10^{-34} J sec

v = Frequency

We can relate the wave and quantum models of electromagnetic radiation behaviour by solving Eq. 1.1 for v and substituting into Eq. 1.2 to obtain (1.3)

$$Q = \frac{hc}{\lambda}$$

Thus, we see that the energy of a quantum is inversely proportional to its wavelength. The longer the wavelength involved, the lower its energy content. This has important implications in remote sensing from the standpoint that naturally emitted long wavelength radiation, such as microwave emission from terrain features, is more difficult to sense than radiation of shorter wavelengths, such as emitted thermal IR energy. The low energy content of long wavelength radiation means that, in general, systems operating at long wavelengths must “view” large areas of the earth at any given time in order to obtain a detectable energy signal.

The sun is the most obvious source of electromagnetic radiation for remote sensing. However, all matter at temperatures above absolute zero (0 K, or - 273^0C) continuously emits electromagnetic radiation. Thus, terrestrial objects are also sources of radiation, although it is of considerably different magnitude and spectral composition than that of the sun. How much energy any object radiates is, among other things, a function of the surface temperature of the object. This property is expressed by the Stefan–Boltzmann law, which states that (1.4)

$$M = \sigma T^4$$

where

M = total radiant exitance from the surface of a material*;* watts ðWÞm_2

σ = *Stefan–Boltzmann constant*, 5. 6697×10^{-8} W m^{-2} K^{-4}

T = absolute temperature (K) of the emitting material

The particular units and the value of the constant are not critical for the student to remember, yet it is important to note that the total energy emitted from an object varies as T4 and therefore increases very rapidly with increases in temperature. Also, it should be noted that this law is expressed for an energy source that behaves as a blackbody. A blackbody is a hypothetical, ideal radiator that totally absorbs and reemits all energy incident upon it. Actual objects only approach this ideal. Suffice it to say for now that the energy emitted from an object is primarily a function of its temperature, as given by Eq. 1.4.

Just as the total energy emitted by an object varies with temperature, the spectral distribution of the emitted energy also varies. Figure 1.4 shows energy distribution curves for blackbodies at temperatures ranging from 200 to 6000 K. The units on the ordinate scale (Wm^{-2} μm^{-1}) express the radiant power coming from a blackbody per 1- μm spectral interval. Hence, the area under these curves equals the total radiant exitance, M, and the curves illustrate graphically what the Stefan–Boltzmann law expresses mathematically: The higher the temperature of the radiator, the greater the total amount of radiation it emits. The curves also show that there is a shift toward shorter wavelengths in the peak of a blackbody radiation distribution as temperature increases. The dominant wavelength, or wavelength at which a blackbody radiation curve reaches a maximum, is related to its temperature by *Wien's displacement law* (1.5)

$$\lambda m = \frac{A}{T}$$

where

λm =wavelength of maximum spectral radiant exitance, μm

A = 2898 μm K

T = temperature, K

Thus, for a blackbody, the wavelength at which the maximum spectral radiant exitance occurs varies inversely with the blackbody's absolute temperature. We observe this phenomenon when a metal body such as a piece of iron is heated. As the object becomes progressively hotter, it begins to glow and its colour changes successively to shorter wavelengths—from dull red to orange to yellow and eventually to white.

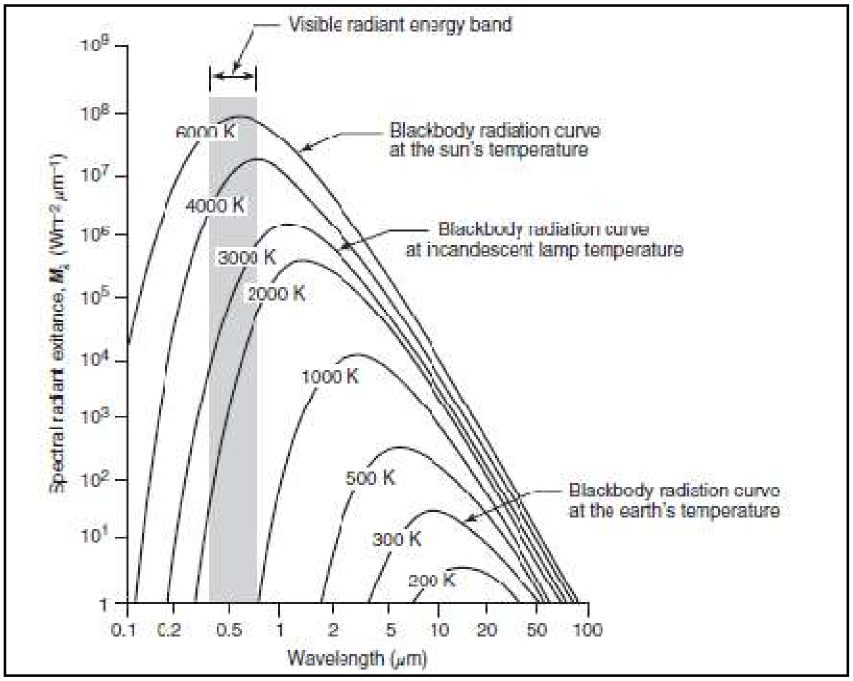


Figure 1.4 Spectral distribution of energy radiated from blackbodies of various temperature.

The sun emits radiation in the same manner as a blackbody radiator whose temperature is about 6000 K (Figure 1.4). Many incandescent lamps emit radiation typified by a 3000 K blackbody radiation curve. Consequently, incandescent lamps have a relatively low output of blue energy, and they do not have the same spectral constituency as sunlight.

The earth's ambient temperature (i.e., the temperature of surface materials such as soil, water, and vegetation) is about 300 K (27°C). From Wien's displacement law, this means the maximum spectral radiant exitance from earth features occurs at a wavelength of about 9.7 μm. Because this radiation correlates with terrestrial heat, it is termed "thermal infrared" energy. This energy can neither be seen nor photographed, but it can be sensed with such thermal devices as radiometers and scanners. By comparison, the sun has a much higher energy peak that occurs at about 0.5 μm, as indicated in Figure 1.4.

Our eyes—and photographic sensors—are sensitive to energy of this magnitude and wavelength. Thus, when the sun is present, we can observe earth features by virtue of reflected solar energy. Once again, the longer wavelength energy emitted by ambient earth features can be observed only with a nonphotographic sensing system. The general dividing line between reflected and emitted IR wavelengths is approximately 3 μm, below this wavelength, reflected energy predominates; above it, emitted energy prevails.

Certain sensors, such as radar systems, supply their own source of energy to illuminate features of interest. These systems are termed "active" systems, in contrast to "passive" systems that sense naturally available energy. A very common example of an active system is a camera utilizing a flash. The same camera used in sunlight becomes a passive sensor.

ENERGY INTERACTIONS IN THE ATMOSPHERE

Irrespective of its source, all radiation detected by remote sensors passes through some distance, or path length, of atmosphere. The path length involved can vary widely. For example, space photography results from sunlight that passes through the full thickness of the earth's atmosphere twice on its journey from source to sensor. On the other hand, an airborne thermal sensor detects energy emitted directly from objects on the earth, so a single, relatively short atmospheric path length is involved. The net effect of the atmosphere varies with these differences in path length and also varies with the magnitude of the energy signal being sensed, the atmospheric conditions present, and the wavelengths involved.

Because of the varied nature of atmospheric effects, we treat this subject on a sensor-by-sensor basis in other chapters. Here, we merely wish to introduce the notion that the atmosphere can have a profound effect on, among other things, the intensity and spectral composition of radiation available to any sensing system. These effects are caused principally through the mechanisms of atmospheric scattering and absorption.

Scattering

Atmospheric scattering is the unpredictable diffusion of radiation by particles in the atmosphere. Rayleigh scatter is common when radiation interacts with atmospheric molecules and other tiny particles that are much smaller in diameter than the wavelength of the interacting radiation. The effect of Rayleigh scatter is inversely proportional to the fourth power of wavelength. Hence, there is a much stronger tendency for short wavelengths to be scattered by this mechanism than long wavelengths.

A "blue" sky is a manifestation of Rayleigh scatter. In the absence of scatter, the sky would appear black. But, as sunlight interacts with the earth's atmosphere, it scatters the shorter (blue) wavelengths more dominantly than the other visible wavelengths. Consequently, we see a blue sky. At sunrise and sunset, however, the sun's rays travel through a longer atmospheric path length than during midday. With the longer path, the scatter (and absorption) of short wavelengths is so complete that we see only the less scattered, longer wavelengths of orange and red.

Rayleigh scatter is one of the primary causes of "haze" in imagery. Visually, haze diminishes the "crispness," or "contrast," of an image. In colour photography, it results in a bluish-gray cast to an image, particularly when taken from high altitude. As we see, haze can often be eliminated or at least minimized by introducing, in front of the camera lens, a filter that does not transmit short wavelengths.

Another type of scatter is Mie scatter, which exists when atmospheric particle diameters essentially equal the wavelengths of the energy being sensed. Water vapour and dust are major causes of Mie scatter. This type of scatter tends to influence longer wavelengths compared to Rayleigh scatter. Although Rayleigh scatter tends to dominate under most atmospheric conditions, Mie scatter is significant in slightly overcast ones.

A more bothersome phenomenon is nonselective scatter, which comes about when the diameters of the particles causing scatter are much larger than the wavelengths of the energy being sensed. Water droplets, for example, cause such scatter. They commonly have a diameter in the range 5 to 100 _m and scatter all visible and near- to mid-IR wavelengths about equally. Consequently, this scattering is "nonselective" with respect to wavelength. In the visible wavelengths, equal quantities of blue, green, and red light are scattered; hence fog and clouds appear white.

Absorption

In contrast to scatter, atmospheric absorption results in the effective loss of energy to atmospheric constituents. This normally involves absorption of energy at a given wavelength. The most efficient absorbers of solar radiation in this regard are water vapour, carbon dioxide, and ozone. Because these gases tend to absorb electromagnetic energy in specific wavelength bands, they strongly influence the design of any remote sensing system. The wavelength ranges in which the atmosphere is particularly transmissive of energy are referred to as atmosphericwindows.

Figure 1.5 shows the interrelationship between energy sources and atmospheric absorption characteristics. Figure 1.5a shows the spectral distribution of the energy emitted by the sun and by earth features. These two curves represent the most common sources of energy used in remote sensing. In Figure 1.5b, spectral regions in which the atmosphere blocks energy are shaded. Remote sensing data acquisition is limited to the nonblocked spectral regions, the atmospheric windows. Note in Figure 1.5c that the spectral sensitivity range of the eye (the "visible" range) coincides with both an atmospheric window and the peak level of energy from the sun. Emitted "heat" energy from the earth, shown by the small curve in (a), is sensed through the windows at 3 to 5 µm and 8 to 14 µm using such devices as thermal sensors. Multispectral sensors observe simultaneously through multiple, narrow wavelength ranges that can be located at various points in the visible through the thermal spectral region. Radar and passive microwavesystems operate through a window in the region 1 mm to 1 m. The important point to note from Figure 1.5 is the interaction and the interdependencebetween the primary sources of electromagnetic energy, the atmosphericwindows through which source energy may be transmitted to and from earth surfacefeatures, and the spectral sensitivity of the sensors available to detect and record theenergy. One cannot select the sensor to be used in any given remote sensing task arbitrarily; one must instead consider (1) the spectral sensitivity of the sensors available, (2) the presence or absence of atmospheric windows in the spectral range(s) in which one wishes to sense, and (3) the source, magnitude, and spectral composition of the energy available in these ranges. Ultimately, however, the choice of spectral range of the sensor must be based on the manner in which the energy interacts with the features under investigation. It is to this last, very important, element that we now turn our attention.

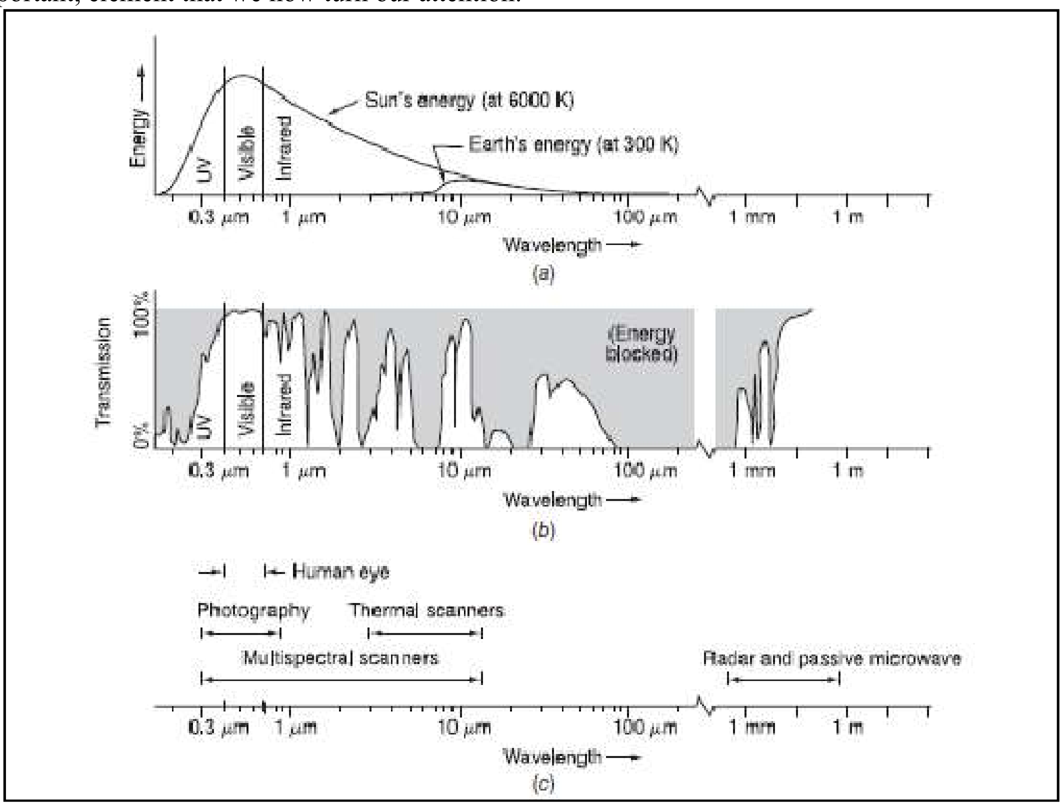


Figure 1.5 Spectral characteristics of (a) energy source (b) atmospheric trasmittance and (c) common remote sensing systems

ENERGY INTERACTIONS WITH EARTH SURFACE FEATURES

When electromagnetic energy is incident on any given earth surface feature, three fundamental energy interactions with the feature are possible. These are illustrated in Figure 1.6 for an element of the volume of a water body. Various fractions of the energy incident on the element are reflected, absorbed, and/or transmitted. Applying the principle of conservation of energy, we can state the interrelationship among these three energy interactions as (1.6)

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda)$$

where

E_I = incident energy
E_R = reflected energy
E_A = absorbed energy
E_T = transmitted energy

with all energy components being a function of wavelength λ.

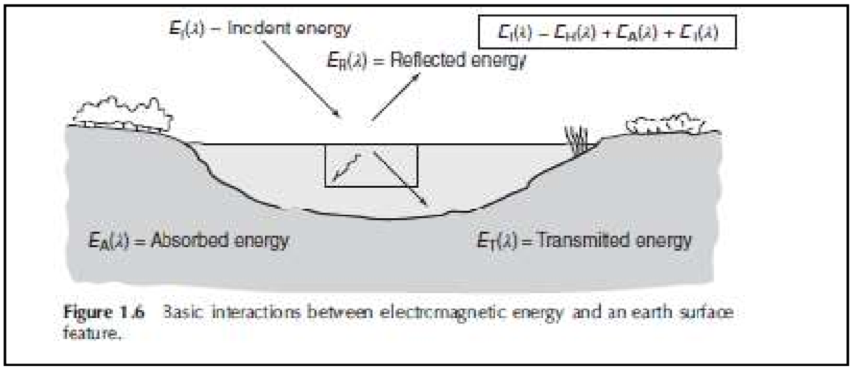


Figure 1.6 Basic interaction between electromagnetic energy and an earth surface feature

Equation 1.6 is an energy balance equation expressing the interrelationship among the mechanisms of reflection, absorption, and transmission. Two points concerning this relationship should be noted. First, the proportions of energy reflected, absorbed, and transmitted will vary for different earth features, depending on their material type and condition. These differences permit us to distinguish different features on an image. Second, the wavelength dependency means that, even within a given feature type, the proportion of reflected, absorbed, and transmitted energy will vary at different wavelengths. Thus, two features may be indistinguishable in one spectral range and be very different in another wavelength band. Within the visible portion of the spectrum, these spectral variations result in the visual effect called colour. For example, we call objects "blue" when they reflect more highly in the blue portion of the spectrum, "green" when they reflect more highly in the green spectral region, and so on. Thus, the eye utilizes spectral variations in the magnitude of reflected energy to discriminate between various objects. Because many remote sensing systems operate in the wavelength regions in which reflected energy predominates, the reflectance properties of earth features are very important. Hence, it is often useful to think of the energy balance relationship expressed by Eq. 1.6 in the form (1.7)

$$E_R(\lambda) = E_I(\lambda) - [E_A(\lambda) + E_T(\lambda)]$$

That is, the reflected energy is equal to the energy incident on a given feature reduced by the energy that is either absorbed or transmitted by that feature. The reflectance characteristics of earth surface features may be quantified by measuring the portion of incident energy that is reflected. This is measured as a function of wavelength and is called spectral reflectance, rl. It is mathematically defined as (1.8)

$$\rho\lambda = \frac{ER\ (\lambda)}{EI\ (\lambda)}$$

$$= \frac{enery\ of\ wavelength\ \lambda\ reflected\ from\ the\ object}{energy\ of\ wavelength\ \lambda\ incident\ upon\ the\ object} \times 100$$

where $\rho\lambda$ is expressed as a percentage.

A graph of the spectral reflectance of an object as a function of wavelength is termed a spectral reflectance curve. The configuration of spectral reflectance curves gives us insight into the spectral characteristics of an object and has a strong influence on the choice of wavelength region(s) in which remote sensing data are acquired for a particular application. This is illustrated in Figure 1.7, which shows highly generalized spectral reflectance curves for deciduous versus coniferous trees. Note that the curve for each of these object types is plotted as a "ribbon" (or "envelope") of values, not as a single line. This is because spectral reflectances vary somewhat within a given material class. That is, the spectral reflectance of one deciduous tree species and another will never be identical, nor will the spectral reflectance of trees of the same species be exactly equal. We elaborate upon the variability of spectral reflectance curves later in this section.

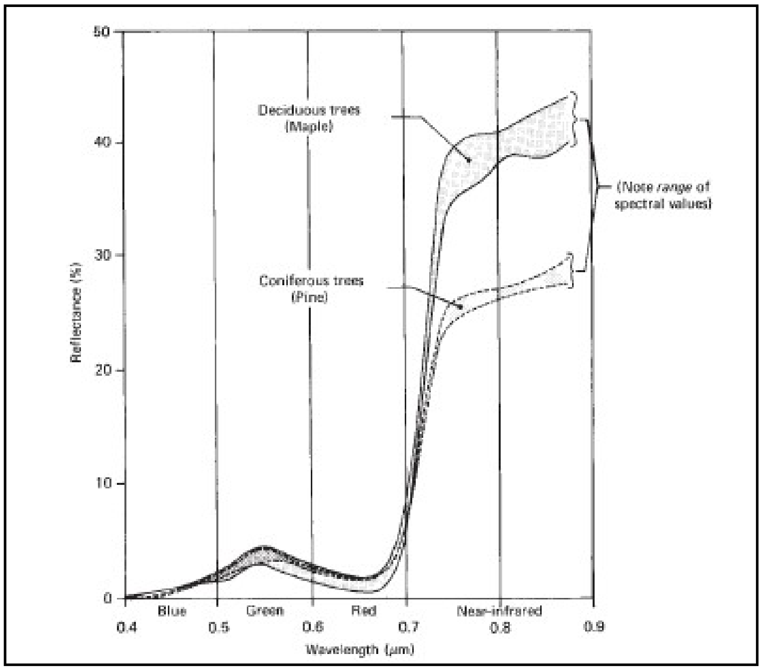


Figure 1.7 Generalized spectral reflectance envelops for deciduous and coniferous trees

In Figure 1.7, assume that you are given the task of selecting an airborne sensor system to assist in preparing a map of a forested area differentiating deciduous versus coniferous trees. One choice of sensor might be the human eye. However, there is a potential problem with this choice. The spectral reflectance curves for each tree type overlap in most of the visible portion of the spectrum and are very close where they do not overlap. Hence, the eye might see both tree types as being essentially the same shade of "green" and might confuse the identity of the deciduous and coniferous trees. Certainly one could improve things somewhat by using spatial clues to each tree type's identity, such as size, shape, site, and so forth. However, this is often difficult to do from the air, particularly when tree types are intermixed. How might we discriminate the two types on the basis of their spectral characteristics alone? We could do this by using a sensor that records near-IR energy. A specialized digital camera whose detectors are sensitive to near-IR wavelengths is just such a system, as is an analog camera loaded with black and white IR film. On near-IR images, deciduous trees (having higher IR reflectance than conifers) generally appear much lighter in tone than do conifers. This is illustrated in Figure 1.8, which shows stands of coniferous trees surrounded by deciduous trees. In Figure 1.8a (visible spectrum), it is virtually impossible to distinguish between tree types, even though the conifers have a distinctive conical shape whereas the deciduous trees have rounded crowns. In Figure 1.8b (near IR), the coniferous trees have a distinctly darker tone. On such an image, the task of delineating deciduous versus coniferous trees becomes almost trivial. In fact, if we were to use a computer to analyze digital data collected from this type of sensor, we might "automate" our entire mapping task. Many remote sensing data analysis schemes attempt to do just that. For these schemes to be successful, the materials to be differentiated must be spectrally separable.

Experience has shown that many earth surface features of interest can be identified, mapped, and studied on the basis of their spectral characteristics. Experience has also shown that some features of interest cannot be spectrally separated. Thus, to utilize remote sensing data effectively, one must know and understand the spectral characteristics of the particular features under investigation in any given application. Likewise, one must know what factors influence these characteristics.

Spectral Reflectance of Earth Surface Feature Types

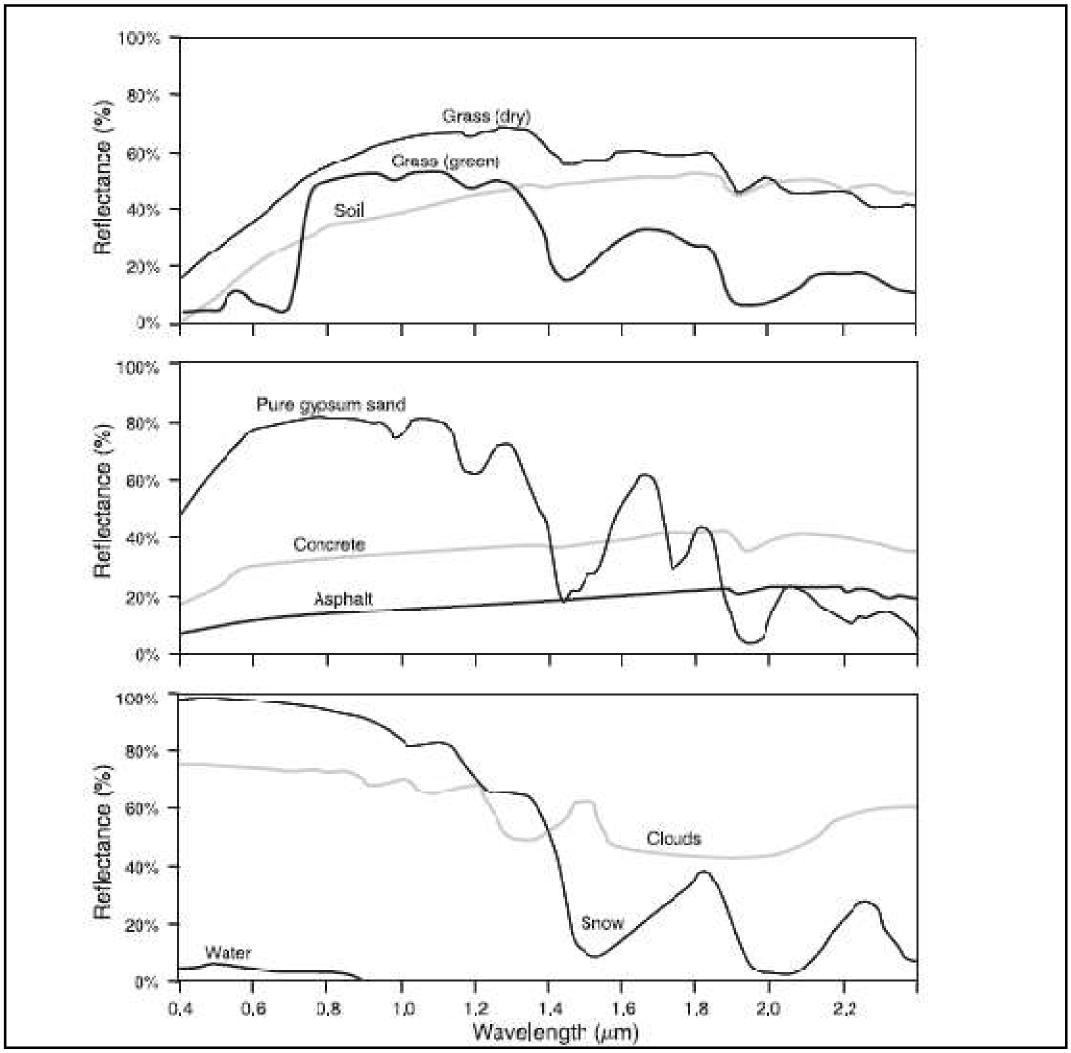


Figure 1.9 shows typical spectral reflectance curves for many different types of features: healthy green grass, dry (non-photosynthetically active) grass, bare soil (brown to dark-brown sandy loam), pure gypsum dune sand, asphalt, construction concrete (Portland cement concrete), fine-grained snow, clouds, and clear lake water. The lines in this figure represent average reflectance curves compiled by measuring a large sample of features, or in some cases representative reflectance measurements from a single typical example of the feature class. Note how distinctive the curves are for each feature. In general, the configuration of these curves is an indicator of the type and condition of the features to which they apply. Although the reflectance of individual features can vary considerably above and

below the lines shown here, these curves demonstrate some fundamental points concerning spectral reflectance.

For example, spectral reflectance curves for healthy green vegetation almost always manifest the "peak-and-valley" configuration illustrated by green grass in Figure 1.9. The valleys in the visible portion of the spectrum are dictated by the pigments in plant leaves. Chlorophyll, for example, strongly absorbs energy in the wavelength bands centered at about 0.45 and 0.67 μm (often called the "chlorophyll absorption bands"). Hence, our eyes perceive healthy vegetation as green in colour because of the very high absorption of blue and red energy by plant leaves and the relatively high reflection of green energy. If a plant is subject to some form of stress that interrupts its normal growth and productivity, it may decrease or cease chlorophyll production. The result is less chlorophyll absorption in the blue and red bands. Often, the red reflectance increases to the point that we see the plant turn yellow (combination of green and red). This can be seen in the spectral curve for dried grass in Figure 1.9.

As we go from the visible to the near-IR portion of the spectrum, the reflectance of healthy vegetation increases dramatically. This spectral feature, known as the red edge, typically occurs between 0.68 and 0.75 μm, with the exact position depending on the species and condition. Beyond this edge, from about 0.75 to 1.3 μm (representing most of the near-IR range), a plant leaf typically reflects 40 to 50% of the energy incident upon it. Most of the remaining energy is transmitted, because absorption in this spectral region is minimal (less than 5%). Plant reflectance from 0.75 to 1.3 μm results primarily from the internal structure of plant leaves. Because the position of the red edge and the magnitude of the near-IR reflectance beyond the red edge are highly variable among plant species, reflectance measurements in these ranges often permit us to discriminate between species, even if they look the same in visible wavelengths. Likewise, many plant stresses alter the reflectance in the red edge and the near-IR region, and sensors operating in these ranges are often used for vegetation stress detection. Also, multiple layers of leaves in a plant canopy provide the opportunity for multiple transmissions and reflections. Hence, the near-IR reflectance increases with the number of layers of leaves in a canopy, with the maximum reflectance achieved at about eight leaf layers (Bauer et al., 1986).

Beyond 1.3 μm, energy incident upon vegetation is essentially absorbed or reflected, with little to no transmittance of energy. Dips in reflectance occur at 1.4, 1.9, and 2.7 μm because water in the leaf absorbs strongly at these wavelengths. Accordingly, wavelengths in these spectral regions are referred to as water absorption bands. Reflectance peaks occur at about 1.6 and 2.2 μm, between the absorption bands. Throughout the range beyond 1.3 μm, leaf reflectance is approximately inversely related to the total water present in a leaf. This total is a function of both the moisture content and the thickness of a leaf.

The soil curve in Figure 1.9 shows considerably less peak-and-valley variation in reflectance. That is, the factors that influence soil reflectance act over less specific spectral bands. Some of the factors affecting soil reflectance are moisture content, organic matter content, soil texture (proportion of sand, silt, and clay), surface roughness, and presence of iron oxide. These factors are complex, variable, and interrelated. For example, the presence of moisture in soil will decrease its reflectance. As with vegetation, this effect is greatest in the water absorption bands at about 1.4, 1.9, and 2.7 μm (clay soils also have hydroxyl absorption bands at about 1.4 and 2.2 μm). Soil moisture content is strongly related to the soil texture: Coarse, sandy soils are usually well drained, resulting in low moisture content and relatively high reflectance; poorly drained fine-textured soils will generally have lower reflectance. Thus, the reflectance properties of a soil are consistent only within particular ranges of conditions. Two other factors that reduce soil reflectance are surface roughness and content of organic matter. The presence of iron oxide in a soil will also significantly decrease reflectance, at least in the visible wavelengths. In any case, it is essential that the analyst be familiar with the conditions at hand. Finally, because soils are essentially opaque to visible and infrared radiation, it should be noted that soil reflectance comes from the uppermost layer of the soil and may not be indicative of the properties of the bulk of the soil.

Sand can have wide variation in its spectral reflectance pattern. The curve shown in Figure 1.9 is from a dune in New Mexico and consists of roughly 99% gypsum with trace amounts of quartz (Jet Propulsion Laboratory, 1999). Its absorption and reflectance features are essentially identical to those of its parent material, gypsum. Sand derived from other sources, with differing mineral compositions, would have a spectral reflectance curve indicative of its parent material. Other factors affecting the spectral response from sand include the presence or absence of water and of organic matter. Sandy soil is subject to the same considerations listed in the discussion of soil reflectance.

As shown in Figure 1.9, the spectral reflectance curves for asphalt and Portland cement concrete are much flatter than those of the materials discussed thus far. Overall, Portland cement concrete tends to be relatively brighter than asphalt, both in the visible spectrum and at longer wavelengths. It is important to note that the reflectance of these materials may be modified by the presence of paint, soot, water, or other substances. Also, as materials age, their spectral reflectance patterns may change. For example, the reflectance of many types of asphaltic concrete may increase, particularly in the visible spectrum, as their surface ages.

In general, snow reflects strongly in the visible and near infrared, and absorbs more energy at mid-infrared wavelengths. However, the reflectance of snow is affected by its grain size, liquid water content, and presence or absence of other materials in or on the snow surface (Dozier and Painter, 2004). Larger grains of snow absorb more energy, particularly at wavelengths longer than $0.8\ \mu m$. At temperatures near 0°C, liquid water within the snowpack can cause grains to stick together in clusters, thus increasing the effective grain size and decreasing the reflectance at near-infrared and longer wavelengths. When particles of contaminants such as dust or soot are deposited on snow, they can significantly reduce the surface's reflectance in the visible spectrum. The aforementioned absorption of mid-infrared wavelengths by snow can permit the differentiation between snow and clouds. While both feature types appear bright in the visible and near infrared, clouds have significantly higher reflectance than snow at wavelengths longer than $1.4\ \mu m$. Meteorologists can also use both spectral and bidirectional reflectance patterns to identify a variety of cloud properties, including ice/water composition and particle size.

Considering the spectral reflectance of water, probably the most distinctive characteristic is the energy absorption at near-IR wavelengths and beyond. In short, water absorbs energy in these wavelengths whether we are talking about water features per se (such as lakes and streams) or water contained in vegetation or soil. Locating and delineating water bodies with remote sensing data are done most easily in near-IR wavelengths because of this absorption property. However, various conditions of water bodies manifest themselves primarily in visible wavelengths. The energy–matter interactions at these wavelengths are very complex and depend on a number of interrelated factors. For example, the reflectance from a water body can stem from an interaction with the water's surface (specular reflection), with material suspended in the water, or with the bottom of the depression containing the water body. Even with deep water where bottom effects are negligible, the reflectance properties of a water body are a function of not only the water per se but also the material in the water.

Clear water absorbs relatively little energy having wavelengths less than about $0.6\ \mu m$. High transmittance typifies these wavelengths with a maximum in the blue-green portion of the spectrum. However, as the turbidity of water changes (because of the presence of organic or inorganic materials), transmittance— and therefore reflectance—changes dramatically. For example, waters containing large quantities of suspended sediments resulting from soil erosion normally have much higher visible reflectance than other "clear" waters in the same geographic area. Likewise, the reflectance of water changes with the chlorophyll concentration involved. Increases in chlorophyll concentration tend to decrease water reflectance in blue wavelengths and increase it in green wavelengths. These changes have been used to monitor the presence and estimate the concentration of algae via remote sensing data. Reflectance data have also been used to determine the presence or absence of tannin dyes from bog vegetation in lowland areas and to detect a number of pollutants, such as oil and certain industrial wastes.

Figure 1.10 illustrates some of these effects, using spectra from three lakes with different bio-optical properties. The first spectrum is from a clear, oligotrophic lake with a chlorophyll level of 1.2 $\mu g/l$ and only 2.4 mg/l of dissolved organic carbon (DOC). Its spectral reflectance is relatively high in the blue-green portion of the spectrum and decreases in the red and near infrared. In contrast, the spectrum from a lake experiencing an algae bloom, with much higher chlorophyll concentration (12.3 $\mu g\ /l$), shows a reflectance peak in the green spectrum and absorption in the blue and red regions. These reflectance and absorption features are associated with several pigments present in algae. Finally, the third spectrum in Figure 1.10 was acquired on an ombrotrophic bog lake, with very high levels of DOC (20.7 mg/l). These naturally occurring tannins and other complex organic molecules give the lake a very dark appearance, with its reflectance curve nearly flat across the visible spectrum.

Many important water characteristics, such as dissolved oxygen concentration, pH, and salt concentration, cannot be observed directly through changes in water reflectance. However, such parameters sometimes correlate with observed reflectance. In short, there are many complex interrelationships between the spectral reflectance of water and particular characteristics. One must use appropriate reference data to correctly interpret reflectance measurements made over water.

Our discussion of the spectral characteristics of vegetation, soil, and water has been very general. The student interested in pursuing details on this subject, as well as factors influencing these characteristics, is encouraged to consult the various references contained in the Works Cited section located at the end of this book.

Spectral Response Patterns

Having looked at the spectral reflectance characteristics of vegetation, soil, sand, concrete, asphalt, snow, clouds, and water, we should recognize that these broad feature types are often spectrally separable. However, the degree of separation between types varies among and within spectral regions. For example, water and vegetation might reflect nearly equally in visible wavelengths, yet these features are almost always separable in near-IR wavelengths.

Because spectral responses measured by remote sensors over various features often permit an assessment of the type and/or condition of the features, these responses have often been referred to as spectral signatures. Spectral reflectance and spectral emittance curves (for wavelengths greater than 3.0 μm) are often referred to in this manner. The physical radiation measurements acquired over specific terrain features at various wavelengths are also referred to as the spectral signatures for those features.

Although it is true that many earth surface features manifest very distinctive spectral reflectance and/or emittance characteristics, these characteristics result in spectral "response patterns" rather than in spectral "signatures." The reason for this is that the term signature tends to imply a pattern that is absolute and unique. This is not the case with the spectral patterns observed in the natural world. As we have seen, spectral response patterns measured by remote sensors may be quantitative, but they are not absolute. They may be distinctive, but they are not necessarily unique.

We have already looked at some characteristics of objects that influence their spectral response patterns. Temporal effects and spatial effects can also enter into any given analysis. Temporal effects are any factors that change the spectral characteristics of a feature over time. For example, the spectral characteristics of many species of vegetation are in a nearly continual state of change throughout a growing season. These changes often influence when we might collect sensor data for a particular application.

Spatial effects refer to factors that cause the same types of features (e.g., corn plants) at a given point in time to have different characteristics at different geographic locations. In small-area analysis the geographic locations may be meters apart and spatial effects may be negligible. When analyzing satellite data, the locations may be hundreds of kilometers apart where entirely different soils, climates, and cultivation practices might exist.

Temporal and spatial effects influence virtually all remote sensing operations. These effects normally complicate the issue of analyzing spectral reflectance properties of earth resources. Again, however, temporal and spatial effects might be the keys to gleaning the information sought in an analysis. For example, the process of change detection is premised on the ability to measure temporal effects. An example of this process is detecting the change in suburban development near a metropolitan area by using data obtained on two different dates.

An example of a useful spatial effect is the change in the leaf morphology of trees when they are subjected to some form of stress. For example, when a tree becomes infected with Dutch elm disease, its leaves might begin to cup and curl, changing the reflectance of the tree relative to healthy trees that surround it. So, even though a spatial effect might cause differences in the spectral reflectances of the same type of feature, this effect may be just what is important in a particular application.

Finally, it should be noted that the apparent spectral response from surface features can be influenced by shadows. While an object's spectral reflectance (a ratio of reflected to incident energy, see Eq. 1.8) is not affected by changes in illumination, the absolute amount of energy reflected does depend on illumination conditions. Within a shadow, the total reflected energy is reduced, and the spectral response is shifted toward shorter wavelengths. This occurs because the incident energy within a shadow comes primarily from Rayleigh atmospheric scattering, and as discussed in Section 1.3, such scattering primarily affects short wavelengths. Thus, in visible-wavelength imagery, objects inside shadows will tend to appear both darker and bluer than if they were fully illuminated. This effect can cause problems for automated image classification algorithms; for example, dark shadows of trees on pavement may be misclassified as water.

Atmospheric Influences on Spectral Response Patterns

In addition to being influenced by temporal and spatial effects, spectral response patterns are influenced by the atmosphere. Regrettably, the energy recorded by a sensor is always modified to some extent by the atmosphere between the sensor and the ground. We will indicate the significance of this effect on a sensor-bysensor basis throughout this book. For now, Figure 1.11 provides an initial frame of reference for understanding the nature of atmospheric effects. Shown in this figure is the typical situation encountered when a sensor records reflected solar energy. The atmosphere affects the "brightness," or radiance, recorded over any given point on the ground in two almost contradictory ways. First, it attenuates (reduces) the energy illuminating a ground object (and being reflected from the object). Second, the atmosphere acts as a reflector itself, adding scattered, extraneous path radiance to the signal detected by the sensor. By expressing these two atmospheric effects mathematically, the total radiance recorded by the sensor may be related to the reflectance of the ground object and the incoming radiation or irradiance using the equation (1.9)

$$L_{tot} \frac{\rho ET}{\pi} + L_p$$

where

L_{tot} =total spectral radiance measured by sensor
ρ =reflectance of object
E = irradiance on object; incoming energy
T = transmission of atmosphere
L_p =path radiance; from the atmosphere and not from the object

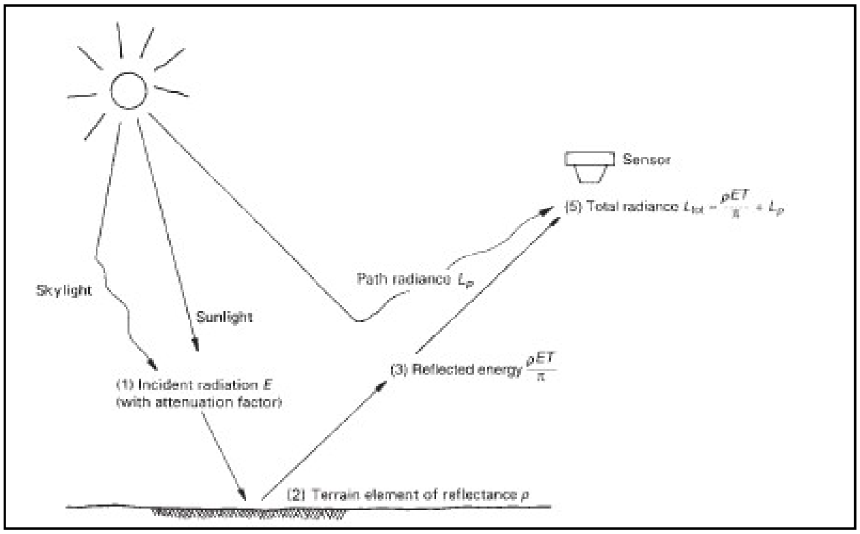


Figure 1.11 Atmospheric effects influencing the measurement of reflected solar energy. Attenuated sunlight and skylight (E) is reflected from a terrain element having reflectance ρ. The attenuated radiance reflected from the terrain element ($\frac{\rho ET}{\pi}$) combines with the path radiance (L_p) to form total radiance (L_{tot}) recorded by the sensor.

It should be noted that all of the above factors depend on wavelength. Also, as shown in Figure 1.11, the irradiance (E) stems from two sources: (1) directly reflected "sunlight" and (2) diffuse "skylight," which is sunlight that has been previously scattered by the atmosphere. The relative dominance of sunlight versus skylight in any given image is strongly dependent on weather conditions (e.g., sunny vs. hazy vs. cloudy). Likewise, irradiance varies with the seasonal changes in solar elevation angle (Figure 7.4) and the changing distance between the earth and sun.

For a sensor positioned close to the earth's surface, the path radiance $L_p$will generally be small or negligible, because the atmospheric path length from the surface to the sensor is too short for much scattering to occur. In contrast, imagery from satellite systems will be more strongly affected by path radiance, due to the longer atmospheric path between the earth's surface and the spacecraft. This can be seen in Figure 1.12, which compares two spectral response patterns from the same area. One "signature" in this figure was collected using a handheld field spectroradiometer, from a distance of only a few cm above the surface. The second curve shown in Figure 1.12 was collected by the Hyperion hyperspectral sensor on the EO-1 satellite. Due to the thickness of the atmosphere between the earth's surface and the satellite's position above the atmosphere, this second spectral response pattern shows an elevated signal at short wavelengths, due to the extraneous path radiance.

In its raw form, this near-surface measurement from the field spectroradiometer could not be directly compared to the measurement from the satellite, because one is observing surface reflectance while the other is observing the so called top of atmosphere (TOA) reflectance. Before such a comparison could be performed, the satellite image would need to go through a process of atmosphericcorrection, in which the raw spectral data are modified to compensate for the expected effects of atmospheric scattering and absorption. This process generally does not produce a perfect representation of the spectral response curve that would actually be observed at the surface itself, but it can produce a sufficiently close approximation to be suitable for many types of analysis.

Geometric Influences on Spectral Response Patterns

The geometric manner in which an object reflects energy is an important consideration. This factor is primarily a function of the surface roughness of the object. Specular reflectors are flat surfaces that manifest mirror-like reflections, where the angle of reflection equals the angle of incidence. Diffuse (or Lambertian) reflectors are rough surfaces that reflect uniformly in all directions. Most earth surfaces are neither perfectly specular nor perfectly diffuse reflectors. Their characteristics are somewhat between the two extremes.

Figure 1.13 illustrates the geometric character of specular, near-specular, near-diffuse, and diffuse reflectors. The category that describes any given surface is dictated by the surface's roughness in comparison to the wavelength of theenergy being sensed. For example, in the relatively long wavelength radio range, a sandy beach can appear smooth to incident energy, whereas in the visible portion of the spectrum, it appears rough. In short, when the wavelength of incident energy is much smaller than the surface height variations or the particle sizes that make up a surface, the reflection from the surface is diffuse.

Diffuse reflections contain spectral information on the "colour" of the reflecting surface, whereas specular reflections generally do not. Hence, in remote sensing,we are most often interested in measuring the diffuse reflectance properties of terrainfeatures.

Because most features are not perfect diffuse reflectors, however, it becomes necessary to consider the viewing and illumination geometry. Figure 1.14 illustrates the relationships that exist among solar elevation, azimuth angle, and viewingangle. Figure 1.15 shows some typical geometric effects that can influence the apparent reflectance in an image. In (a), the effect of differential shading is illustrated in profile view. Because the sides of features may be either sunlit or shaded, variations in brightness can result from identical ground objects at different locations in the image. The sensor receives more energy from the sunlit side of the tree at B than from the shaded side of the tree at A. Differential shading is clearly a function of solar elevation and object height, with a stronger effect at low solar angles. The effect is also compounded by differences in slope and aspect (slope orientation) over terrain of varied relief.

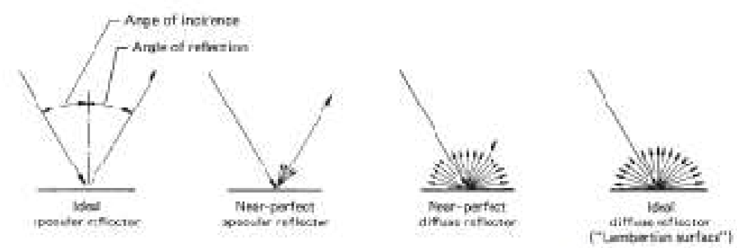


Figure 1.13 Specular versus diffuse reflectance. (We are most often interested in measuring the diffuse reflectance of objects.)

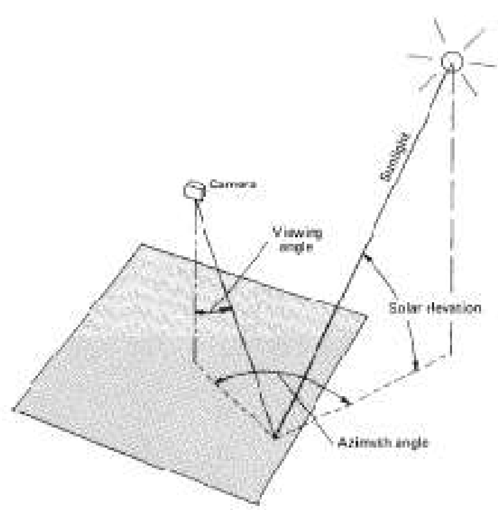


Figure 1.14 Sun object image angular relationship

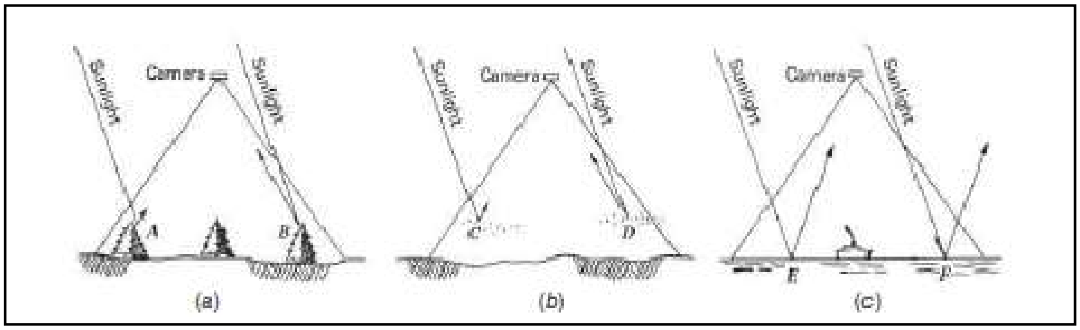


Figure 1.15 (a) differential shading (b) differential scattering (c) specular reflection

Figure 1.15b illustrates the effect of differential atmospheric scattering. As discussed earlier, backscatter from atmospheric molecules and particles adds light (path radiance) to that reflected from ground features. The sensor records more atmospheric backscatter from area D than from area C due to this geometric effect. In some analyses, the variation in this path radiance component is small and can be ignored, particularly at long wavelengths. However, under hazy conditions, differential quantities of path radiance often result in varied illumination across an image.

As mentioned earlier, specular reflections represent the extreme in directional reflectance. When such reflections appear, they can hinder analysis of the imagery. This can often be seen in imagery taken over water bodies. Figure 1.15c illustrates the geometric nature of this problem. Immediately surrounding point E on the image, a considerable increase in brightness would result from specular reflection.

UNIT – 6: DIFFERENT SENSORS AND RESOLUTIONS

The Landsat System

The Landsat earth resources satellite system was the first designed to provide near global coverage of the earth's surface on a regular and predictable basis. The first three Landsats had identical orbit characteristics, as summarised in Table A.4. The orbits were near polar and sun synchronous – i.e., the orbital plane precessed about the earth at the same rate that the sun appears to move across the face of the earth. In this manner data was acquired at about the same local time on every pass.

Table A.4. Landsat 1, 2, 3 orbit characteristics

Orbit:	Sun synchronous, near polar; nominal 9:30 am descending equatorial crossing; inclined at about 99° to the equator
Altitude:	920 km (570 mi)
Period:	103 min
Repeat Cycle:	14 orbits per day over 18 days (251 revolutions)

All satellites acquired image data nominally at 9:30 a.m. local time on a descending (north to south) path; in addition Landsat 3 obtained thermal data on a night-time ascending orbit for the few months that its thermal sensor was operational. Fourteen complete orbits were covered each day, and the fifteenth, at the start of the next day, was 159 km advanced from orbit 1, thus giving a second day coverage contiguous with that of the first day. This advance in daily coverage continued for 18 days and then repeated. Consequently complete coverage of the earth's surface was given, with 251 revolutions in 18 days.

The orbital characteristics of the second generation Landsats, commencing with Landsats 4 and 5, are different from those of their predecessors. Again image data is acquired nominally at 9:30 a.m. local time in a near polar, sun synchronous orbit; however the spacecraft are at the lower altitude of 705 km. This lower orbit gives a repeat cycle of 16 days at 14.56 orbits per day. This corresponds to a total of 233 revolutions every cycle. Table A.5 summarises the Landsat 4, 5 orbit characteristics. Unlike the orbital pattern for the first generation Landsats, the day 2 ground pattern for Landsats 4 and 5 is not adjacent and immediately to the west of the day 1 orbital pattern. Rather it is displaced the equivalent of 7 swath centres to the west. Over 16 days this leads to the repeat cycle. Landsat 6, launched in 1993, was not successfully placed in orbit and was lost over the Atlantic Ocean. Landsat 7 is a similar satellite in all respects. Whereas Landsats 1, 2 and 3 contained on-board tape recorders for temporary storage of image data when the satellites were out of view of earth stations, Landsats 4 and 5 do not, and depend on transmission either to earth stations directly or via the geosynchronous communication satellite TDRS (Tracking and Data Relay Satellite).

TDRS is a high capacity communication satellite that is used to relay data from a number of missions, including the Space Shuttle. Its ground receiving station is in White Sands, New Mexico from which data is relayed via domestic communication satellites. Landsat 7 also uses TDRS for data downlinking but has an on-board solid state recorder for temporary storage.

Orbit:	Near polar, sun synchronous; nominal 9:30 am descending equatorial crossing (10:00 am for Landsat 7)
Altitude:	705 km
Period:	98.9 min
Repeat Cycle:	14.56 orbits per day over 16 days (total of 233 revolutions)

The Landsat Instrument Complement

Three imaging instruments have been used with the Landsat satellites to date. These are the Return BeamVidicon (RBV), the Multispectral Scanner (MSS) and the Thematic Mapper (TM). Table A.6 shows the actual imaging payload for each satellite along with historical data on launch and out-of-service dates. Two different RBV's were used: a multispectral RBV package was incorporated on the first two satellites, while a panchromatic instrument with a higher spatial resolution was used on Landsat 3. The MSS on Landsat 3 also contained a thermal band; however this operated only for a few months. The MSSwas not used after Landsat 5.With the launch of Landsat 7 an Enhanced Thematic Mapper + (ETM+) was added. The following sections provide an overview of the three Landsat instruments, especially from a data characteristic point-of-view.

Table A.6. Landsat payloads, launch and out of service dates

Satellite	Imaging Instruments			Launched	Out-of-service
Landsat 1	RBV^m	MSS		23 Jul 1972	6 Jan 1978
Landsat 2	RBV^m	MSS		22 Jan 1975	27 Jul 1983
Landsat 3	RBV^p	MSS^t		5 Mar 1978	7 Sept 1983
Landsat 4		MSS	TM	16 Jul 1982	Aug 1993
Landsat 5		MSS	TM	1 Mar 1984	–
Landsat 6			ETM	lost on launch	–
Landsat 7			ETM+	15 Apr 1999	–

m – multispectral RBV
p – panchromatic RBV
t – MSS with thermal band

The Return Beam Vidicon (RBV)

As the name suggests the RBV's were essentially television camera-like instruments that took "snapshot" images of the earth's surface along the ground track of the satellite. Image frames of 185 km ×185 km were acquired with each shot, repeated at 25 s intervals to give contiguous frames in the along track direction at the equivalent ground speed of the satellite.

Three RBV cameras were used on Landsats 1 and 2, distinguished by different transmission filters that allowed three spectral bands of data to be recorded as shown in Table A.7. On Landsat 3 two RBV cameras were used; however both operated panchromatically and were focussed to record data swaths of 98 km, overlapped to give a total swath of about 185 km. By so doing a higher spatial resolution of 40 m was possible, by comparison to 80 m for the earlier RBV system. Historically the spectral ranges recorded by the RBV's on Landsats 1 and 2 were referred to as bands 1, 2 and 3. The MSS bands (see following) in the first generation of Landsats were numbered to follow on in this sequence.

The Multispectral Scanner (MSS)

The Multispectral Scanner was the principal sensor on Landsats 1, 2 and 3 and was the same on each spacecraft with the exception of an additional band on Landsat 3. The MSS is a mechanical scanning device that acquires data by scanning the earth's surface in strips normal to the satellite motion. Six lines are swept simultaneously by an oscillating mirror and the reflected solar radiation so monitored is detected in four wavelength bands for Landsats 1 and 2, and five bands for Landsat 3, as shown in Table A.7. A schernatic illustration of the six line scanning pattern used by the MSS is shown in Fig. A.1. It is seen that the sweep pattern gives rise to an MSS swath width of 185 km thereby corresponding to the image width of the RBV.

The width of each scan line corresponds to 79 m on the earth's surface so that the six lines simultaneously correspond to 474 m. Approximately 390 complete six-line scans are collected to provide an effective image that is also 185 km in the along track direction. For Landsats 1 and 2, 24 signal detectors were required to provide four spectral bands from each of the six scan lines. A further two were added for the thermal band data of Landsat 3. Those detectors are illuminated by radiation reflected from the oscillating scanning mirror in the MSS, and produce a continuously varying

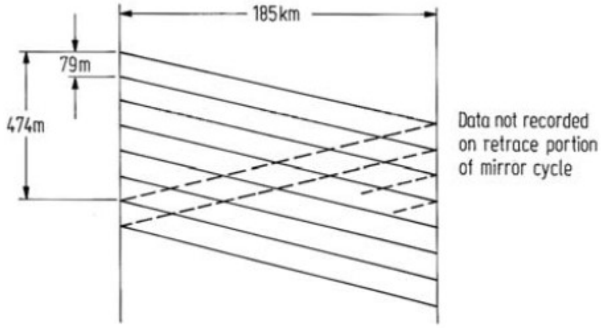


Fig. A.1. The six line scanning pattern used by the Landsat multispectral scanner. Dimensions are in equivalent measurements on the ground. This scanning pattern is the same in each of bands 4 to 7. The same six line pattern is used on Landsats 4 and 5 except that the strip width is 81.5 m and 82.5 m respectively

Table A.7. Characteristics of the Landsat imaging devices

Instrument	Spectral bands (μm)			IFOV (m)	Dynamic range (bits)
RBV[m]	1.	0.475– 0.575	(blue)	79 × 79	
	2.	0.580– 0.680	(red)	79 × 79	
	3	0.689– 0.830	(near IR)	79 × 79	
RBV[p]		0.505– 0.750	(panchromatic)	40 × 40	
MSS	4.[a]	0.5 – 0.6	(green)	79 × 79	7
	5	0.6 – 0.7	(red)	79 × 79	7
	6.	0.7 – 0.8	(near IR)	79 × 79	7
	7.	0.8 – 1.1	(near IR)	79 × 79	6
	8.[b]	10.4 –12.6	(thermal)	237 × 237	
TM	1.	0.45 – 0.52	(blue)	30 × 30	8
	2.	0.52 – 0.60	(green)	30 × 30	8
	3	0.63 – 0.69	(red)	30 × 30	8
	4.	0.76 – 0.90	(near IR)	30 × 30	8
	5.	1.55 – 1.75	(mid IR)	30 × 30	8
	7[c]	2.08 – 2.35	(mid IR)	30 × 30	8
	6.	10.4 –12.5	(thermal)	120 × 120	8
ETM+	1.	0.450– 0.515	(blue)	30 × 30	8
	2.	0.525– 0.605	(green)	30 × 30	8
	3.	0.630– 0.690	(red)	30 × 30	8
	4.	0.775– 0.900	(near IR)	30 × 30	8
	5	1.550– 1.750	(mid IR)	30 × 30	8
	7.	2.090– 2.350	(mid IR)	30 × 30	8
	6.	10.40 –12.50	(thermal)	60 × 60	8
	pan	0.520– 0.900		13 × 15	8

[a] MSS bands 4 to 7 have been renumbered MSS bands 1 to 4 from Landsat 4 onwards. IFOV = 81.5, 82.5 m for Landsats 4, 5.
[b] MSS band 8 was used only on Landsat 3.
[c] TM band 7 is out of sequence since it was added last in the designafter the previous six bands had been firmly established. It was incorporated at the request of the geological community owing to the importance of the 2 μm region in assessing hydrothermal alteration.

electrical signal corresponding to the energy received along the 79mwide associated scan line. The optical aperture of the MSS and its detectors for bands 4 to 7 is such that at any instant of time each detector sees a pixel that is 79 m in size also along the scan line. Consequently the effective pixel size (or instantaneous field of view IFOV) of the detectors is 79 m × 79 m. At a given instant the output from a detector is the integrated response from all cover types present in a 79m×79m region of the earth's surface. Without any further processing the signal from the detector would appear to be varying continuously with time. However it is sampled in time to produce discrete measurements across a scan line. The sampling rate corresponds to pixel centres of 56 m giving a 23 m overlap of the 79 m × 79 m pixels, as depicted in Fig. A.2. The thermal infrared band on Landsat 3, band 8, has an IFOV of 239 m × 239 m. As a result there are only two band 8 scan lines corresponding to the six for bands 4 to 7, as indicated above.

The IFOV's of the multispectral scanners on Landsats 4 and 5 have been modified to 81.5 m and 82.5 m respectively although the pixel centre spacing of 56 m has been retained. In addition the bands have been renamed as bands 1, 2, 3 and 4, corresponding to bands 4, 5, 6 and 7 from the earlier missions. After being spatially sampled, the data from the detectors is digitised in amplitude into 6 bit words. Before so-encoding, the data for bands 4, 5 and 6 is compressed allowing decompression into effective 7 bit words upon reception at a ground station.

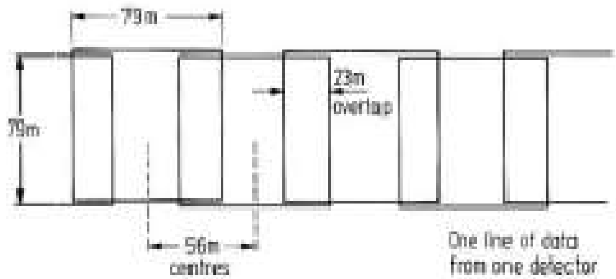


Fig. A.2. The relationship between instantaneous field of view and pixel overlap for Landsat MSS pixels

The Thematic Mapper (TM) and Enhanced Thematic Mapper + (ETM+)

The Thematic Mapper is a mechanical scanning device as for the MSS, but hasimproved spectral, spatial and radiometric characteristics. Seven wavelength bandsare used, with coverage as shown in Table A.7. Note that band 7 is out of place in theprogression of wavelengths, it having been added, after the initial planning phase, atthe request of the geological community. The Enhanced Thematic Mapper + carriedon Landsat 7 includes a panchromatic band and improved spatial resolution on the thermal band.

Whereas the MSS of all Landsats scans and obtains data in one direction only, the TM acquires data in both scan directions again with a swath width of 185 km. Sixteen scan lines are swept simultaneously giving a 480 m strip across the satellite path, as illustrated in Fig. A.3. This permits a lower mirror scan rate compared with the MSS and thus gives a higher effective dwell time for a given spot on the ground, making possible the higher spatial resolution and improved dynamic range.

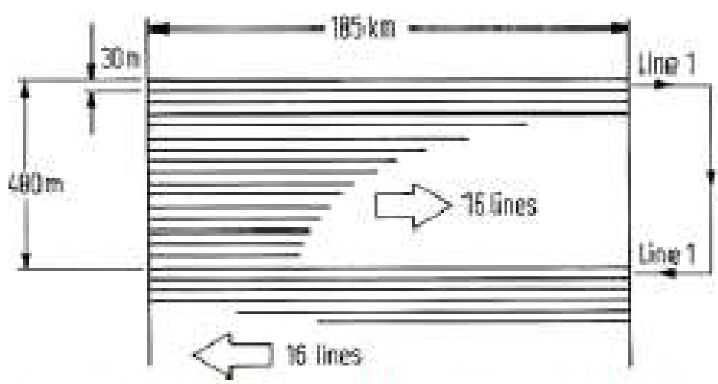


Fig. A.3. Scanning characteristics of the Landsat Thematic Mapper

The SPOT HRV, HRVIR, HRG, HRS and Vegetation Instruments

The early French SPOT satellites (Système pour d'Observation de la Terre), carried two imaging devices referred to as HRV's. These instruments utilize a different technology for image acquisition from that employed in the Landsat MSS and TM devices. Rather than using oscillating mirrors to provide cross-track scanning during the forward motion of the space platform, the SPOT HRV instruments consist of a linear array of charge coupled device (CCD) detectors. These form what is commonly referred to as a "push broom" scanner. Each detector in the array scans a strip in the along track direction. By having several thousand such detectors a wide swath can be imaged without the need for mechanical scanning. Moreover, owing to the long effective dwell time this allows for each pixel, a higher spatial resolution is possible.

A trade-off however is that charge coupled device technology was not available for wavelengths into the middle infrared range at the time of early SPOT development. Consequently the spectral bands provided by the HRV are not unlike those of the Landsat MSS. The HRV covers a ground swath width of 60 km; two instruments are mounted side by side in the spacecraft to give a total swath width of 117 km, there being a 3km overlap of the individual swaths.

Two imaging modes are possible. One is a multispectral mode and the other panchromatic. The imaging characteristics of these are summarised, along with the satellite orbital properties, in Table A.8. An interesting property of the HRV is that it incorporates a steerable mirror to allow imaging to either side of nadir. This allows daily coverage for a short period along with a stereoscopic viewing capability.

SPOT 4, launched in March 1998, carries two instruments – the HRVIR (High Resolution Visible and Infrared) and the Vegetation instrument. Characteristics of both are summarised in Table A.8. SPOT 5 carries an instrument known as HRG (High Resolution Geometry) that uses essentially the same wavebands as the HRVIR but with a higher spatial resolution. Its characteristics are also summarised in Table A.8. SPOT 5 also carries the Vegetation instrument, along with a new device called the HRS (High Resolution Stereoscopy) that images fore and aft of the spacecraft to allow stereoscopic products to be developed.

Table A.8. Spot satellite and sensor characteristics

Orbit:	Near polar, sun synchronous; nominal 10.30 am descending equatorial crossing
Altitude:	832 km
Period:	101 min
Repeat Cycle:	26 days

Satellite	Imaging Instrument	Launched	Out-of-Service
SPOT 1	HRV (×2)	22 Feb 1986 reactivated 9 Jan 1997	5 Jan 1991
SPOT 2	HRV (×2)	22 Jan 1990	
SPOT 3	HRV (×2)	26 Sep 1993	14 Nov 1996
SPOT 4	HRVIR (×2), Vegetation	24 Mar 1998	
SPOT 5	HRG (×2), HRS, Vegetation	4 May 2002	

Instrument	Spectral bands (μm)	IFOV (m)	Swath (km)	Dynamic range (bits)
HRV[m]	0.50 – 0.59 (green)	20 × 20	60	8
	0.61 – 0.68 (red)	20 × 20	60	8
	0.79 – 0.89 (near IR)	20 × 20	60	8
HRV[p]	0.51 – 0.73	10 × 10	60	8
HRVIR[m]	0.50 – 0.59 (green)	20 × 20	60	8
	0.61 – 0.68 (red)	20 × 20	60	8
	0.78 – 0.89 (near IR)	20 × 20	60	8
	1.58 – 1.75 (mid IR)	20 × 20	60	8
HRVIR[p]	0.61 – 0.68	10 × 10	60	8
HRG[m]	0.50 – 0.59 (green)	10 × 10	60	8
	0.61 – 0.68 (red)	10 × 10	60	8
	0.79 – 0.89 (near IR)	10 × 10	60	8
	1.58 – 1.75 (mid IR)	20 × 20	60	8
HRG[p]	0.49 – 0.69	5 × 5 and 2.5 × 2.5	60	8
Vegetation	0.45 – 0.52 (blue)	1000 × 1000	2250	10
	0.61 – 0.68 (red)	1000 × 1000	2250	10
	0.78 – 0.89 (near IR)	1000 × 1000	2250	10
	1.58 – 1.75 (mid IR)	1000 × 1000	2250	10
HRS	0.49 – 0.69	5 × 10	120	8

[m] multispectral mode
[p] panchromatic mode

Stereoscopy) that images fore and aft of the spacecraft to allow stereoscopic products to be developed.

ADEOS (Advanced Earth Observing Satellite)

ADEOS-I was launched by the Japanese space agency NASDA in August 1996. It carried a number of imaging instruments and non-imaging sensors, including OCTS (Ocean Colour and Temperature Sensor), AVNIR (Advanced Visible and Near Infrared Radiometer), NSCAT (NASA Spectrometer), TOMS (Total Ozone Mapping Spectrometer),POLDER(Polarization and Directionality of the Earth's Reflectance), IMG (lnterferometric Monitor for Greenhouse Gases), ILAS (Improved Limb Atmospheric Sensor) and RIS (Retroreflector in Space). Its successor, ADEOS-II, was launched on 14 December 2002. The sensors on ADEOS-II (now also called MIDORI-II) are the GLI (Global Imager), AMSR (Advanced Microwave Scanning Radiometer), ILAS-II, POLDER and SeaWINDS. Characteristics of the OCTS, GLI and AVNIR are given in Table A.9, along with spacecraft orbital details.

Table A.9. ADEOS satellite and sensor characteristics

Orbit:	near polar, sun synchronous; nominal 10.30 am equatorial crossing
Altitude:	797 km
Period:	101 min
Repeat Cycle:	41 days (ADEOS 2 has a 4 day repeat cycle)

Instrument	Spectral bands (μm)	IFOV (m)	Swath (km)	Dynamic range (bits)
AVNIR[m]	0.42 – 0.50	16 × 16	80	8
	0.52 – 0.60	16 × 16	80	8
	0.61 – 0.69	16 × 16	80	8
	0.76 – 0.89	16 × 16	80	8
AVNIR[p]	0.52 – 0.69	8 × 8	80	7
OCTS	0.412 ± 0.01	700 × 700*	1400	10
	0.443 ± 0.01	700 × 700	1400	10
	0.490 ± 0.01	700 × 700	1400	10
	0.520 ± 0.01	700 × 700	1400	10
	0.565 ± 0.01	700 × 700	1400	10
	0.670 ± 0.01	700 × 700	1400	10
	0.765 ± 0.02	700 × 700	1400	10
	0.865 ± 0.02	700 × 700	1400	10
	3.55 – 3.88	700 × 700	1400	10
	8.25 – 8.80	700 × 700	1400	10
	10.3 – 11.4	700 × 700	1400	10
	11.4 – 12.7	700 × 700	1400	10
GLI	0.375 12.5 (36 bands @ 10 nm bandwidth)	1000 × 100 or 250 × 250	1600	12

Sea-Viewing Wide Field of View Sensor (SeaWiFS)

In August 1997 the OrbView-2 (SeaStar) satellite was launched, carrying the Sea- WiFS sensor with characteristics as shown in Table A.10. Its wavebands have been chosen with ocean-related applications in mind.

Table A.10. SeaStar satellite and SeaWiFS sensor characteristics

Orbit:	near polar, sun synchronous			
Altitude:	705 km			
Period:	98.9 min			
Repeat Cycle:	1 day			

Instrument	Spectral bands (μm)	IFOV (m)	Swath (km)	Dynamic range (bits)
SeaWiFS	0.402 – 0.422	1100 × 1100	2800	10
	0.433 – 0.453	1100 × 1100	2800	10
	0.480 – 0.500	1100 × 1100	2800	10
	0.500 – 0.520	1100 × 1100	2800	10
	0.545 – 0.565	1100 × 1100	2800	10
	0.660 – 0.680	1100 × 1100	2800	10
	0.745 – 0.785	1100 × 1100	2800	10
	0.845 – 0.885	1100 × 1100	2800	10

Table A.11. MOS orbit and sensor characteristics

MOS:		Altitude	908 km
		Orbit	sun synchronous, 99.1° inclination 10–11 am equatorial crossing
		Repeat Cycle	17 days
MESSR:	Bands	0.51–0.59 μm 0.61–0.69 μm 0.73–0.80 μm 0.80–1.10 μm	
	IFOV	50 m × 50 m	
	Dynamic range	8 bit	
	Swath per MESSR	100 km	
VTIR:	Bands	0.5– 0.7 μm 6.0– 7.0 μm 10.5–11.5 μm 11.5–12.5 μm	
	IFOV	900 m × 900 m for visible channel 2700 m × 2700 m for the others	
	Dynamic range	8 bit	
	Swath Width	1500 km	

Marine Observation Satellite (MOS)

The Marine Observation Satellites MOS-l and MOS-lb were launched by Japan in February 1987 and February 1990 respectively and were taken out of service in March 1995 and April 1996 respectively. While intended largely for oceanographic studies, the data from the satellites' two optical imaging sensors – the MESSR (Multispectrum Electronic Self Scanning Radiometer) and the VTIR (Visible and Thermal Infrared Radiometer) are of value to land based remote sensing as well. The satellites also carried a Microwave Scanning Radiometer (MSR) intended for water vapour, snow and ice studies. Two MESSRs were used to provide side by side observations. Each has a 100 km swath width; with an overlap in coverage of 15 km the total available swath is 185 km. Orbital details of the MOS satellites and characteristics of their optical sensors are given in Table A.11.

A.2.10

Indian Remote Sensing Satellite (IRS)

A series of remote sensing satellites has been launched by India since March 1988. They carry imaging systems known as the LISS (Linear Imaging Self Scanner), the WiFS (Wide Field Sensor), the advanced version AWiFS, the Ocean Colour Monitor (OCM), the Multifrequency Scanning Microwave Radiometer (MSMR), the Molecular Optoelectronic Scanner (MOS) and a panchromatic sensor. Orbital details of the satellites and characteristics of the sensors are summarised in Table A.12.

A.2.11

RESURS-O1

Russia has orbited a series of remote sensing satellites since 1985 under the name RESURS-O1. TableA.13 gives platform and sensor characteristics for the third in the series. The principal sensor, from which commercially available imagery is produced, is the MSU-SK, which is a conically scanning instrument.

The Earth Observing 1 (EO-1) Mission

EO-1 was launched on 21 November 2000 into the same orbit as Landsat 7, but one minute behind, allowing near simultaneous, partly overlapping coverage. The Terra platform (see Sect. A.2.13) is essentially also in the same orbit, but 30 minutes behind EO-1. The two imaging instruments of importance on EO-1 are the Advanced Land Imager (ALI) and Hyperion, the characteristics of which are given in Table A.14.

Table A.12. IRS satellite and sensor characteristics

Orbit:	Near polar, sun synchronous; nominal 10.35 am equatorial crossing
Altitude:	904 km (1A, 1B), 817 km (1C), 736/825 km (1D)
Period:	101 min
Repeat Cycle:	22 days (1A, 1B), 24 days (1C, 1D)

Satellite	Imaging Instrument	Launched	Out-of-Service
IRS-1A	LISSI	Mar 1988	
IRS-1B	LISSII	Aug1991	
IRS-P2	LISSII	16 Oct 1994	
IRS-1C	LISSIII, WiFS		
IRS-1D	LISSIII, Pan WiFS	29 Sep 1997	
IRS-P3	WiFS, MOS	21 Mar 1996	
IRS-P4	OCM, MSMR	May 1999	(Oceansat 1)
IRS-P5	Pan	Scheduled 2004	(Cartosat 1)
IRS-P6	LISSIII, LISS IV, AWiFS	17 Oct 2003	(Resourcesat 1)

Instrument	Spectral bands (μm)	IFOV (m)	Swath (km)	Dynamic range (bits)
LISSI, II	0.45 – 0.52	73 × 73 (LISSI)	146	7
	0.52 – 0.59	36 × 36 (LISSII)	146	7
	0.62 – 0.68		146	7
	0.77 – 0.86		146	7
LISSIII	0.52 – 0.59	23 × 23	142 – 146	7
	0.62 – 0.68	23 × 23	142 – 146	7
	0.77 – 0.86	23 × 23	142 – 146	7
	1.55 – 0.59	70 × 70	142 – 146	7
Pan	0.5 – 0.57	10 × 10	70	7
WiFS	0.62 – 0.68	188 × 188	774	7
	0.77 – 0.86	188 × 188	774	7
LISSIV	0.53 – 0.59	5.8 × 5.8	23.9 (XS mode)	7
	0.62 – 0.68	5.8 × 5.8	70.3 (Pan mode)	7
	0.77 – 0.86	5.8 × 5.8		7
AWiFS	0.52 – 0.59	56 × 56	740	10
	0.62 – 0.68	56 × 56	740	10
	0.77 – 0.86	56 × 56	740	10
	1.55–1.70	56 × 56	740	10
OCM	0.4 – 0.885 (8 bands @ 20 nm bandwidth)	360 × 360	1420	12
MOS-A	0.756 – 0.768 (4 bands @ 1.4 nm bandwidth)	1570 × 1400	195	16
-B	0.408 – 1.015 (13 bands @ 10 mm bandwidth)	520 × 520	200	16
-C	1.600 (1 band @ 100 mm bandwidth)	520 × 640	192	16

Table A.15. Aqua and Terra sensor characteristics

Instrument	Spectral Bands (µm)	IFOV (m)	Swath (km)	Dynamic Range (bits)
MODIS*	0.620–0.670	250 × 250	2330	12
	0.841–0.876	250 × 250	2330	12
	0.459–2.155 (5 bands)	500 × 500	2330	12
	0.405–14.385 (29 bands)	1000 × 1000	2330	12
ASTER	0.52–0.60	15 × 15	60	8
	0.63–0.69	15 × 15	60	8
	0.75–0.86	15 × 15	60	8
	0.75–0.86 (backward looking)	15 × 15	60	8
	1.600–1.700	30 × 30	60	8
	2.145–2.185	30 × 30	60	8
	2.185–2.225	30 × 30	60	8
	2.235–2.285	30 × 30	60	8
	2.295–2.365	30 × 30	60	8
	2.360–2.430	30 × 30	60	8
	8.125–8.475	90 × 90	60	12
	8.475–8.825	90 × 90	60	12
	8.925–9.275	90 × 90	60	12
	10.250–10.950	90 × 90	60	12
	10.950–11.650	90 × 90	60	12

* The band description for MODIS is quite complex, since groups of bands are targeted on specific applications. Full details can be obtained from the MODIS home page at http://modis.gsfc.nasa.gov

Aqua and Terra

The Aqua and Terra platforms are part of NASA's Earth Observing System. They were launched respectively on 4 May 2002 and 18 December 1999 in sun synchronous orbits comparable to those for Landsat 7, but with descending equatorial crossings of 1:30 am (or 1:30 pm ascending) for Aqua and around 10:30 am for Terra. They are also known as the Earth Observing System (EOS) PM (Aqua) and EOS AM platforms (Terra).

The principal instruments on Terra are MODIS (Moderate Resolution Imaging Spectrometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Spectrometer), CERES (Clouds and the Earth's Radiant Energy System), MISR (Multi-angle Imaging Spectro Radiometer) and MOPITT (Measurement of Pollution in the Troposphere). The instrument complement on Aqua includes MODIS, a set of optical and microwave atmospheric sounders, CERES, and a scanning microwave radiometer. The characteristics of MODIS and ASTER are given in Table A.15.

Ikonos

The Ikonos satellite was launched on 24 September 1999. It is in a sun-synchronous, near polar orbit at an altitude of 681 km, with a repeat cycle of 35 days (although a 1.5 day revisit capacity is possible with off-nadir pointing). It has a descending equatorial crossing of 10:30 am, comparable to SPOT. Characteristics of its imaging sensor are given in Table A.16.

Table A.16. Ikonos sensor characteristics

	Spectral Bands (µm)	IFOV (m)	Swath (km)	Dynamic Range (bits)
Panchromatic	0.45–0.90	1 × 1	11	11
Blue	0.45–0.53	4 × 4	11	11
Green	0.52–0.61	4 × 4	11	11
Red	0.64–0.72	4 × 4	11	11
Near IR	0.77–0.88	4 × 4	11	11

Aircraft Scanners in the Visible and Infrared Regions
General Considerations

Multispectral line scanners, similar in principle to the Landsat MSS and TM instruments, have been available for use in civil aircraft since the late 1960's and early 1970's. As with satellite image acquisition it is the forward motion of the aircraft that provides along track scanning whereas a rotating mirror or a linear detector array provides sensing in the across track direction. There are several operational features that distinguish the data provided by aircraft scanners from that produced by satellite-borne devices. These are of significance to the image processing task. First, the data volume can be substantially higher. This is a result of having (i) a large number of spectral bands or channels available and (ii) a large number of pixels produced per mission, owing to the high spatial resolution available. Frequently up to 1000 pixels may be recorded across the swath, with many thousands of scan lines making up a flight line; each pixel is normally encoded to at least 8 bits.

A second feature of importance relates to field of view (FOV) – that is the scan angle either side of nadir over which data is recorded. This is depicted in Fig. A.4. In the case of aircraft scanning the FOV, 2γ , is typically about 70 to 90◦. Such a large angle is necessary to acquire an acceptable swath of data from aircraft altitudes. By comparison the FOV for the Landsats 1 to 3 is 11.56◦ while that for Landsats 4 and 5 is slightly larger at about 15◦. The consequence of the larger FOV with aircraft scanning is that significant distortions in image geometry can occur at the edges of the scan. Often these have to be corrected by digital processing. Finally, the attitude stability of an aircraft as a remote sensing platform is much poorer than the stability of a satellite in orbit, particularly the Landsat 4 generation for which the pointing accuracy is 0.01◦ with a stability of 10−6 degrees per second. Because of atmospheric turbulence, variations in aircraft attitude described by pitch, roll and yaw can lead to excessive image distortion. Sometimes the aircraft scanner is mounted on a three axis stabilized platform to minimise these variations. It is more common however to have the scanner fixed with respect to the aircraft body and utilize a variable sampling window on the data stream to compensate for aircraft roll.

Use of airborne multispectral scanners offers a number of benefits. Often the user can select the wavebands of interest in a particular application, and small bandwidths can be used. Also, the mission can be flown to specific user requirements concerning time of day, bearing angle and spatial resolution, the last being established by the aircraft height above ground level. As against these however, data acquisition from aircraft platforms is expensive by comparison with satellite recording since aircraft missions are generally flown for a single user and do not benefit from the volume market and synoptic view available to satellite data.

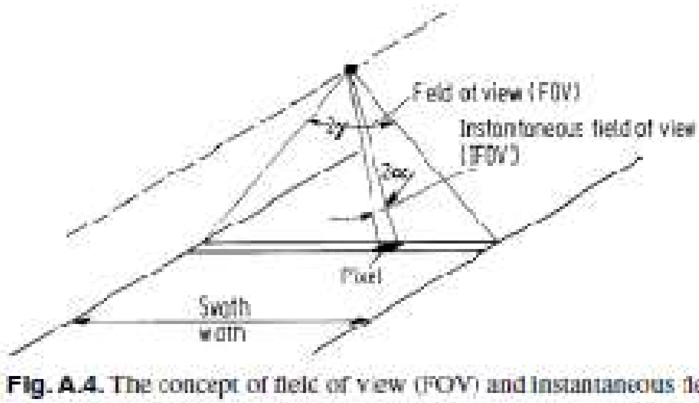


Fig. A.4. The concept of field of view (FOV) and instantaneous field of view (IFOV)

Airborne Imaging Spectrometers

Since the mid 1980's the availability of new detector technologies has made possible the development of aircraft scanners capable of recording image data in a large number, typically hundreds, of spectral channels. For a given pixel, enough samples of its reflectance properties may be obtained by these instruments to allow very accurate characterisation of the pixel's spectral reflectance curve over the visible and reflected infrared region. Because of the large number of channels the data sets are often referred to as hyperspectral. These devices were the forerunners of instruments such as Hyperion, treated in Sect. A.2.12. A good discussion of the development of imaging spectrometry may be found in Goetz et al. (1985) and Vane and Goetz (1988). Much of the work on those devices led to the development of similar spaceborne instruments. Table A.17 summarises the characteristics of a selection of current aircraft imaging spectrometers.

Spaceborne Imaging Radar Systems

The Seasat SAR

The first earth observational space mission to carry a synthetic aperture imaging radar was the Seasat satellite launched in June 1978. Although only short lived it recorded about 126 million square kilometres of image data, including multiple coverage of many regions. Several other remote sensing instruments were also carried, including a radar altimeter, a scatterometer, a microwave radiometer and a visible and infrared imaging radiometer. Relevant characteristics of the satellite and its SAR are summarised in Table A.18. Polarization referred to in this table relates to the orientation of the electric field vector in the transmitted and received waveforms.

Free space propagation of electromagnetic energy, such as that used for radar, takes place as a wave with electric and magnetic field vectors normal to each other and also normal to the direction of propagation. Should the electric field vector be parallel to the earth's surface, the wave is said to be horizontally polarized. Should it be vertical then the wave is said to be vertically polarized. A wavefront with a combination of the two will be either elliptically or circularly polarized. Even though one particular polarization might be adopted for transmission, some rotation can occur when the energy is reflected from the ground. Consequently at the receiver often both vertically and horizontally polarized components are available, each having its own diagnostic properties concerning the earth cover type being sensed. Whether one or the other, or both, are received depends upon the antenna used with the radar. In the case of Seasat, horizontally polarized radiation was transmitted (***H***) and horizontally polarized returns were received (***H***). Further details on the Seasat SAR will be found in Elachi et al. (1982).

Spaceborne (Shuttle) Imaging Radar-A (SIR-A)

A modified version of the Seasat SAR was flown as the SIR-A sensor on the second flight of Space Shuttle in November of 1981. Although the mission was shortened to three days, image data of about 10 million square kilometres was recorded. In contrast to Seasat however, in which the final image data was available digitally, the data in SIR-A was recorded and processed optically and thus is available only in film format. For digital processing therefore it is necessary to have areas of interest digitized from film using a device such as a scanning micro densitometer. A summary of SIR-A characteristics is given in Table A.18, wherein it will be seen that the incidence angle was chosen quite different from that for the Seasat SAR. Interesting features of landform can be brought out by processing the two together. More details on SIR-A will be found in Elachi et al. (1982) and E1achi (1983).

Table A.17. Imaging spectrometers

Instrument	Spectral range µm	Spectral resolution nm	Dynamic range bits	IFOV mrad	Pixels per line
CASI-2 (Itres Research)	0.4 – 1 (288 channels)	2.2	12	1.3	512
CASI-3 (Itres Research)	0.4 – 1.05 (288 channels)	2.2	14	1.3	1490
DAIS 7915 (Geophysical Environmental Research Corp.)	0.4 – 1.0 (32 channels)	15 – 30	15	3.3	512
	1.5 – 1.8 (8 channels)	45			
	2 - 2.5 (32 channels)	20			
	3 – 5 (1 channel)	2000			
	8 – 12.6 (6 channels)	900			
AVIRIS (Airborne Visible and Infrared Imaging Spectrometer – JPL)	0.4 – 0.72 (31 channels)	9.7	12	1	550
	0.69 – 1.30 (63 channels)	9.6			
	1.25 – 1.87 (63 channels)	8.8			
	1.84 – 2.45 (63 channels)	11.6			
MIVIS (Daedalus Enterprises Inc.)	0.433 - 0.833 (20 channels)	20	12	2	765
	1.15 – 1.55 (8 channels)	50			
	2.00 – 2.50 (64 channels)				
	8.20 – 12.70 (10 channels)	≤500			
HYDICE (Hyperspectral Digital Image Collection Experiment US Naval Research Labs)	0.4 – 2.5 (206 channels)	7.6 – 14.9	12	0.5	320
HYMAP (Integrated Spectronics Pty Ltd)	0.44 – 0.88	16	12	2.5 × 2.0	512
	0.881 – 1.335	13			
	1.4 – 1.81	12			
	1.95 – 2.5 (128 bands total)	16			

Table A.18. Characteristics of Seasat SAR, SIR-A, SIR-B, and SIR-C

	Seasat	SIR-A	SIR-B	SIR-C/X-SAR
Altitude	800 km	245 km	225–235 km	225 km
Wavelength	0.235 m	0.235 m	0.235 m	L – 0.235 m, C – 0.058 m, X – 0.031 m
Polarization	*HH*	*HH*	*HH*	HH, HV, VH, VV (only VV at X)
Incidence angle	20°	47°	15–57°	20–55°
Swath width	100 km	50 km	20–50 km	15–90 km
Range resolution	25 m	40 m	58–17 m	13–26 m (L, C), 10–20 (X)
Azimuth resolution	25 m	40 m	25 m	30 m

Spaceborne (Shuttle) Imaging Radar-B (SIR-B)

SIR-B, the second instrument in the NASA shuttle imaging radar program was carried on Space Shuttle mission 41G in October 1984. Again the instrument was essentially the same as that used on Seasat and SIR-A, however the antenna was made mechanically steerable so that the incidence angle could be varied during the mission. Also about half the data was recorded digitally with the remainder being optically recorded. Details of the SIR-B mission are summarised in Table A.18; NASA (1984) contains further information on the instrument and experiments planned for the mission. Because of the variable incidence angle both the range resolution and swath width also varied accordingly.

Spaceborne (Shuttle) Imaging Radar-C (SIR-C)/X-Band Synthetic Aperture Radar (X-SAR)

SIR-C/X-SAR, the third Shuttle radar mission, was carried out over two 10 day flights in April and September 1994. The SAR carried was the result of cooperation between NASA and DARA, the German Aerospace Agency, and had the characteristics indicated in Table A.18. Further details of the mission and the SAR will be found in Stofan et al. (1995) and Jordan et al. (1995).

ERS-1,2

The European Remote Sensing Satellites ERS-1 and ERS-2 were launched in July 1991 and April 1995 respectively; they carry a number of sensors, one of which is synthetic aperture radar intended largely for sea state and oceanographic applications. Characteristics of the radar are summarised in Table A.19.

Table A.19. Characteristics of free flying satellite SAR systems

	ERS-1, 2	JERS-1	Radarsat
Altitude	785 km	568 km	793–821 km
Wavelength*	0.057 m	0.235 m	0.057 m
Polarization	*VV*	*HH*	*HH*
Incidence angle	23°	35°	19–49°
Swath width	100 km	75 km	50–500 km
Range resolution	30 m	18 m	27 m
Azimuth resolution	30 m	18 m	19–24 m

* 5.3 GHz for ERS 1, 2 and Radarsat and 1.28 Ghz for JERS-1

JERS-1

The Japanese Earth Resources Satellite JERS-1 was launched in February 1992. It carries two imaging instruments; one is an optical sensor and the other imaging radar. Table A.19 shows the design characteristics for the radar. The optical sensor, called OPS, has 8 wavebands between 0.52 µm and 2.40 µm with a swath width of 75 km and a dynamic range of 6 bits. The optical pixel size is 18.3 m (across track) × 24.2 m (along track). Provision is included for stereoscopic imaging.

Radarsat

Canada's Radarsat was launched on 4 November 1995; its SAR is able to operate in the standard and six non-standard modes, one of which will give a 518 km swath (Raney et al., 1991). Table A.19 lists the characteristics of the Radarsat SAR. Radarsat-2, scheduled for launch in 2005, will have an ultra-fine beam mode with 3 m resolution, and further polarisation options (VV, VH, HV).

Shuttle Radar Topography Mission (SRTM)

By deploying an outboard radar antenna 60 m from the space shuttle, along with the main antenna in the cargo bay, two simultaneous images can be obtained of the same region, but from different perspectives. Because of the coherent nature of the data, the two images can be interfered to reveal topographic detail of the earth's surface. The Shuttle Radar Topography Mission in February 2000 used the SIR-C (C band)/X-SAR system to acquire interferometric data from which approximately 80% of the earth's land mass was imaged with 16 m absolute height accuracy and 20 m horizontal accuracy.

Envisat Advanced Synthetic Aperture Radar (ASAR)

The ASAR is an advanced version of the synthetic aperture radar from the ERS-1 and 2 missions. It operates at 5.331GHzand incorporates a number of imaging modes that provide a variety of resolutions, polarisations and swath widths. Generally, the swath width is 100 km with the exception of wave mode (5 km) and wide swath width and global monitoring (400 km) products.

The Advanced Land Observing Satellite (ALOS) PALSAR

ALOS is scheduled for launch in 2005, and is designed as a follow on to JERS-1 and ADEOS (Midori). Besides PRISM (for stereoscopic mapping) and an AVNIR ALOS will carry a phased array L band SAR, to be known as PALSAR. The SAR will have a swath width of 70 km and a 2 look spatial resolution of 10 m in its observation mode, and a swath width of 250–360 km with a spatial resolution of 100 m in a scansar (wide swath width) mode.

Table A.20. Representative aircraft synthetic aperture radar systems

	CCRS[1] SAR	DLR[2] ESAR	JPL[3] AIRSAR
Wavebands	X, C bands	X, C, L & P bands	C, L & P bands
Polarisation	HH, VV	multipolarisation	multipolarisation
Range resolution	6 m, 20 m	1.5 m	10 m
Azimuth resolution	6 m, 10 m	4–12 m	1 m
Interferometry	yes	yes	yes

[1] Canada Centre for Remote Sensing
[2] Deutsche Forschungsanstalt für Luft- und Raumfahrt
[3] Jet Propulsion Laboratory

Aircraft Imaging Radar Systems

Airborne imaging radar systems in SLAR and SAR technologies are also available. As with airborne multispectral scanners these offer a number of advantages over equivalent satellite based systems including flexibility in establishing mission parameters (bearing, incidence angle, spatial resolution etc.) and proprietary rights to data. However the cost of data acquisition is also high. TableA.20 summarises the characteristics of three aircraft imaging radars, chosen to illustrate the operating parameters of these devices by comparison to satellite based systems. Note the band: wavelength designations – X: 0.030 m, C: 0.057 m, L: 0.235 m, P: 0.667 m. Note also that interferometric operation is also possible with the systems listed.

UNIT -7: DIGITAL IMAGE PROCESSING – PRINCIPLES AND APPROACHES

Digital image analysis refers to the manipulation of digital images with the aid of a computer. This could range from an amateur photographer using freely available software to adjust the contrast and brightness of pictures from her or his digital camera to a team of scientists using neural-network classification to map mineral types in an airborne hyperspectral image. In this chapter, we will begin with simple and widely used methods for enhancing digital images, correcting errors, and generally improving image quality prior to further visual interpretation or digital analysis. Many of these techniques are broadly applicable to a wide range of types of remotely sensed data. We then proceed to cover more advanced and specialized techniques. The enhancement, processing, and analysis of digital images comprise an extremely broad subject, often involving procedures that can be mathematically complex. For most topics in this chapter, we focus on the concepts and principles involved, without delving too deeply into the mathematics and algorithms that are used to implement these digital analysis methods. The use of computers for digital processing and analysis began in the 1960s with early studies of airborne multispectral scanner data and digitized aerial photographs. However, it was not until the launch of Landsat-1 in 1972 that digital image data became widely available for land remote sensing applications. At that time, not only was the theory and practice of digital image processing in its infancy, but also the cost of digital computers was very high, and their computational efficiency was very low by modern standards. Today, access to low-cost, efficient computer hardware and software is commonplace, and the sources of digital image data are many and varied. These sources range from commercial and governmental earth resource satellite systems, to the meteorological satellites, to airborne scanner data, to airborne digital camera data, to image data generated by photogrammetric scanners and other high resolution digitizing systems. All of these forms of data can be processed and analyzed using the techniques described in this chapter.

The central idea behind digital image processing is quite simple. One or more images are loaded into a computer. The computer is programmed to perform calculations using an equation, or series of equations, that take pixel values from the raw image as input. In most cases, the output will be a new digital image whose pixel values are the result of those calculations. This output image may be displayed or recorded in pictorial format or may itself be further manipulated by additional software. The possible forms of digital image manipulation are seemingly infinite. However, virtually all these procedures may be categorized into one (or more) of the following seven broad types of computer-assisted operations -

Image preprocessing. These operations aim to correct distorted or degraded image data to create a more faithful representation of the original scene and to improve an image's utility for further manipulation later on. This typically involves the initial processing of raw image data to eliminate noise present in the data, to calibrate the data radiometrically, to correct for geometric distortions, and to expand or contract the extent of an image via mosaicking or subsetting. These procedures are often termed preprocessingoperations because they normally precede further manipulation and analysis of the image data to extract specific information.

Image enhancement. These procedures are applied to image data in order to more effectively render the data for subsequent interpretation. In many cases, image enhancement involves techniques for heightening the visual distinctions among features in a scene, ultimately increasing the amount of information that can be interpreted from the data.

Image classification. The objective of image classification is to replace visual interpretation of image data with quantitative techniques for automating the identification of features in a scene. This normally involves the analysis of multiple bands of image data (typically multispectral, multitemporal, polarimetric, or other sources of complementary information) and the application of statistically based decision rules for determining the land cover identity of each pixel in an image. When these decision rules are based solely on the spectral radiances observed in the data, we refer to the classification process as spectral pattern recognition. In contrast, the decision rules may be based on the geometric shapes, sizes, and patterns present in the image data. These procedures fall into the domain of spatial pattern recognition. Hybrid methods, in which both spatial and spectral patterns are used for classification, are increasingly common. In any case, the intent of the classification process is to categorize all pixels in a digital image into one of several land cover classes, or "themes." These categorized data may then be used to produce thematic maps of the land cover present in an image and/or produce summary statistics on the areas covered by each land cover type. We emphasize "supervised,""unsupervised," and "hybrid" approaches to spectrally based image classification before introducing more specialized topics such as object based classification, classification of mixed pixels, and the use of neural networks in the classification process. Finally, we describe various procedures for assessing and reporting the accuracy of image classification results.

Analysis of change over time. Many remote sensing projects involve the analysis of two or more images from different points in time, to determine the extent and nature of changes over time. With the proliferation of temporally rich data sets that offer dozens or hundreds of images of a given area over a period of weeks or years, there is also a new interest in time-series analysis of remotely sensed imagery, such as the analysis of seasonal cycles, interannual variability, and long-term trends in dense multitemporal data sets.

Data fusion and GIS integration. These procedures are used to combine image data for a given geographic area with other geographically referenced data sets for the same area. These other data sets might simply consist of image data generated on other dates by the same sensor or by other remote sensing systems. Frequently, the intent of data merging is to combine remotely sensed data with other ancillary sources of information in the context of a GIS. For example, image data are often combined with soil, topographic, ownership, zoning, and assessment information.

Hyperspectral image analysis. Virtually all of the image processing principles introduced in this chapter in the context of multispectral image analysis may be extended directly to the analysis of hyperspectral data. However, the basic nature and sheer volume of hyperspectral data sets is such that various image processing procedures have been developed to analyze such data specifically.

7. Biophysical modeling. The objective of biophysical modeling is to relate quantitatively the digital data recorded by a remote sensing system to biophysical features and phenomena measured on the ground. For example, remotely sensed data might be used to estimate such varied parameters as crop yield, pollution concentration, or water depth.

Data Types and Data Exploration

Introduction

The geographic information system (GIS) is merely a tool. Much like a saw is useless without wood, or a screwdriver useless without a screw, the GIS is useless without geospatial data. Geospatial data can be acquired from any number of sources such as printed maps, geospatial databases, or data acquired with the Global Positioning System (GPS).

Geospatial data is generally grouped into two types: raster and vector. For the data to be most useful, it must be organized logically. In this module, you will examine various data models and the recommended methods for storing and organizing geospatial data. You will also experiment with some basic data exploration techniques in ArcGIS.

Organizing Geospatial Data

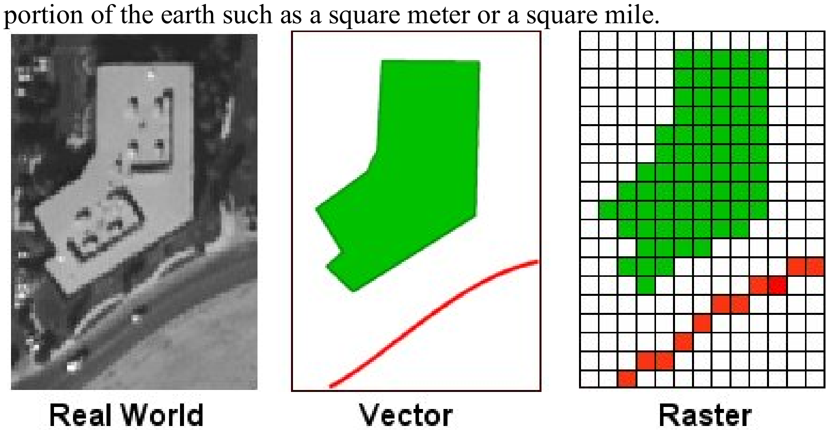
Geospatial data includes information about the location of an object or phenomenon as well as information about that object or phenomenon. Everything on earth has a location, so there is almost no limit to the amount of geospatial information that can be captured and stored.

Because GIS users will frequently be dealing with large amounts of data, it is important to store and organize geospatial data in a logical folder structure or geodatabase. Once the data is stored and organized logically, GIS users can quickly and easily find the data they need and new files can be added to related files in the appropriate location. After you have created a logical structure, it is best to avoid constantly moving data around within it.

Geographic Data Models

All GIS data attempts to define abstract real-world features in a data model, or format that can be understood by a computer and ArcGIS. In the GIS world, there are two main data models used to represent features: the vector data model and the raster data model.

The vector data model represents discrete objects on the surface of the earth—such as trees, rivers, or lakes—as point, line, and polygon features with well-defined boundaries. A raster data model represents the surface of the earth as a grid of equally sized cells. An individual cell represents a portion of the earth such as a square meter or a square mile.



The vector data model represents geographic features with exactly defined boundaries, while the raster data model represents them as cells of the same value. Notice that the shapes of the raster building and road do not seem as similar to the real-world shapes as the vector shapes do.

Both the vector and raster data models are useful for representing geospatial data, but one may be more appropriate than the other when it comes to representing a particular type of geospatial data or answering different kinds of questions. In general, the vector data model is useful for representing features that have discrete boundaries, while the raster data model is most useful for representing continuous geospatial data—phenomena such as elevation, precipitation, and temperature—which do not have well-defined boundaries and which usually change gradually across a given area.

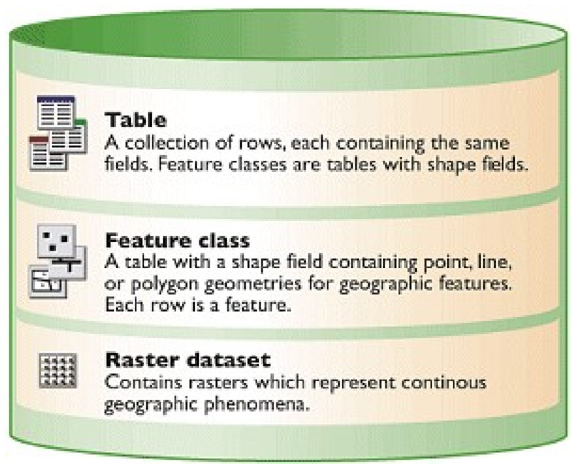
Later in this module, you will learn more about these two data models and some common geographic data formats that are based on each.

Geodatabases

To assemble geospatial data into a GIS database, you could create a collection of folders containing data stored in different formats, such as shapefiles and coverages, or you could create a geodatabase.

The geodatabase is a data storage format introduced by Esri with ArcGIS® software. In technical terms, the geodatabase is a relational database, a type of database composed of various tables that organize data and are linked to one another.

It may be more helpful at this point, however, to just think of a geodatabase as a container for storing geographic data. The geographic data stored in a geodatabase may be a collection of vector feature classes (point, line, polygon, or annotation), raster datasets, and tables.



The basic geodatabase components are tables, feature classes, and raster datasets.

Organizing data into a geodatabase has certain advantages: in a geodatabase all your data is stored in one central location instead of being spread out in different files. Especially when you have a large amount of data, having all your related data stored in one geodatabase helps you maintain an overview of your data holdings and more easily find the data you need when you need it.

The geodatabase also promotes faster and more accurate data entry and editing than other formats. For example, you can set up rules for a geodatabase feature class that say only certain values are valid for a particular attribute, and you can create relationships among feature classes so that when a feature in one feature class is updated, related features in other feature classes update as well.

Types of Geodatabases

There are three types of geodatabases:

- File geodatabases

- Personal geodatabases for Microsoft® Access™
- ArcSDEgeodatabases

The type of geodatabase you create depends on what your data will be used for as well as the structure and workflows of your organization.

If a geodatabase will be used by a relatively small workgroup, and the data will be edited by a single user, a file geodatabase is most suitable. A file geodatabase, which has a GDB extension, is portable across operating systems and can handle very large datasets with very fast performance. Its storage capacity is virtually unlimited and it requires less disk space than other file formats. The file geodatabase is the recommended data format for ArcGIS. You create the file geodatabase with ArcCatalog.

The personal geodatabase for Access is also designed for small workgroups with a single editor. It uses the Microsoft Access data format with the MDB file extension and supports Microsoft Access table operations for working with attribute values. The personal geodatabase for Access has a 2 gigabyte (GB) size limit and is only supported on the Microsoft Windows operating system. You create the personal geodatabase for Access with ArcCatalog.

Large organizations with significant amounts of data that needs to be accessed and edited by multiple people would likely consider an ArcSDEgeodatabase the best choice. All ArcSDEgeodatabases require a relational database management system such as DB2, Oracle, or SQL Server and Esri'sArcSDE® technology, which is included with ArcGIS® Server software.

File geodatabases, personal geodatabases for Access, and ArcSDEgeodatabases store the same basic elements—feature classes (stand-alone or in feature datasets), raster datasets, and nonspatial tables.

Note: Unlike shapefiles, geodatabases cannot be read by non-ArcGIS software. Feature classes stored in geodatabases, however, can be exported to file formats that can be read by non-ArcGIS software.

Types of Geospatial Data Files

ArcGIS supports geospatial datasets that are managed in geodatabases as well as in numerous standalone GIS file formats. Geodatabase datasets represent the native data structure for ArcGIS and are the primary data format used for editing and data management. Yet, many additional datasets can be used.

A number of additional file formats are supported. Shapefiles, for example, are a common data format that you will probably come across on the Internet or when working with ArcPad® software or global positioning system (GPS) applications. File formats like these can be used in ArcGIS much like geodatabase datasets—to create layers in ArcMap™ and ArcGlobe™; as inputs for geoprocessing operations; to be viewed and queried in charts, maps, globes, and tables; and converted to and from many other GIS formats.

Viewing Geospatial Data

In ArcGIS, you can view geospatial data in either ArcCatalog™ or ArcMap.

ArcCatalog is an Esri application for managing spatial data. The interface is similar to Windows Explorer and is used to copy, move, and delete geospatial data. In addition, ArcCatalog is used to obtain information about spatial data. ArcCatalog also assists in such tasks as the creation of metadata, or "data about data." You will learn more about metadata later in this module.

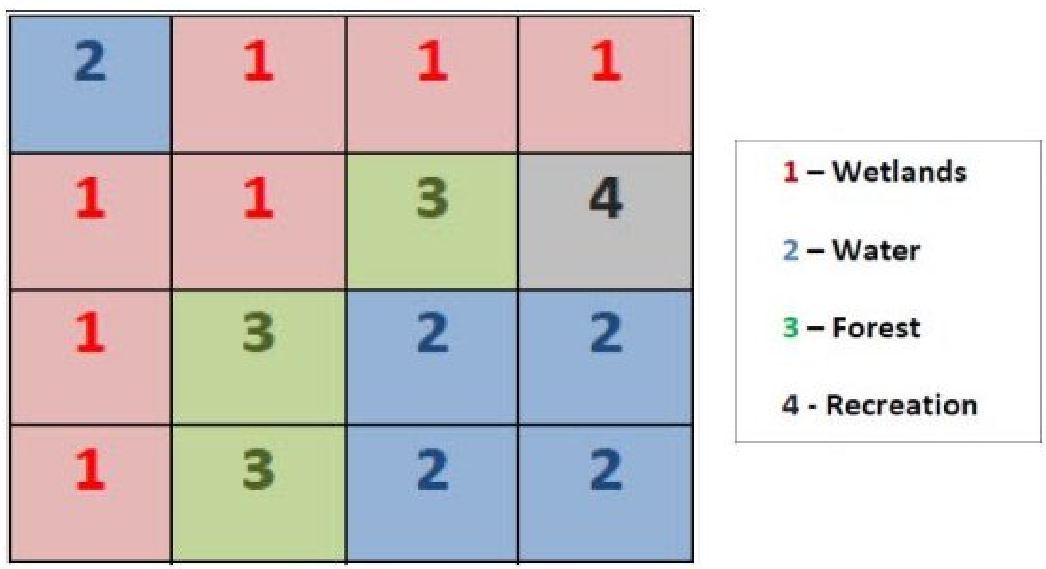
ArcMap is the central Esri application used in ArcGIS. ArcMap is where you display and explore the datasets for your study area, where you assign symbols, and where you create map layouts for printing or publication. ArcMap is also the application you use to create and edit datasets. ArcMap represents geographic information as a collection of layers and other elements in a map document, or MXD file.

Raster Data

A raster is made up of a grid of square cells where all the cells are the same size. (Pixels is another term for cells. Short for "picture element," pixels usually refers to image cells.) Thus, an individual raster cell represents a portion of the earth such as a square meter or square mile.

Each cell has a given value, depending on the type of dataset. In general, these cells are squares and are organized in rows and columns to form a rectangular dataset. The number of rows and columns in a raster does not have to be the same. For example, your raster may have 2000 rows and 1857 columns.

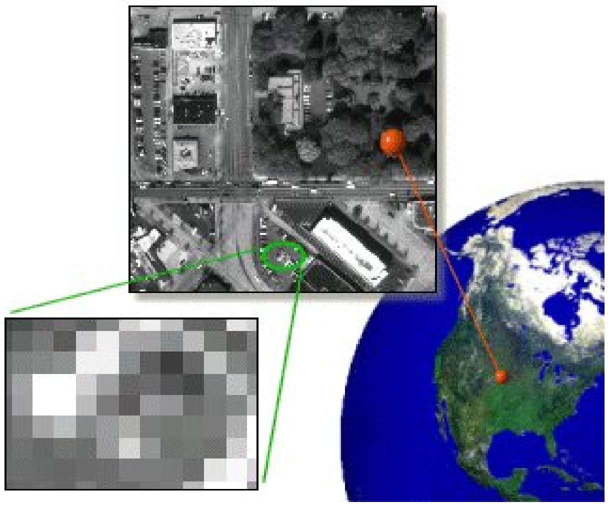
Note: There are exceptions to this, such as raster datasets with rectangular rather than square cells or even datasets comprised of an arrangement of hexagons or triangles. However, most of the raster datasets you will encounter will have the standard square cell.



The cell is the basic spatial unit of a raster. Each cell has a numeric value. Colors or shades are used to display the different cell values, yielding an image or a map of some kind.

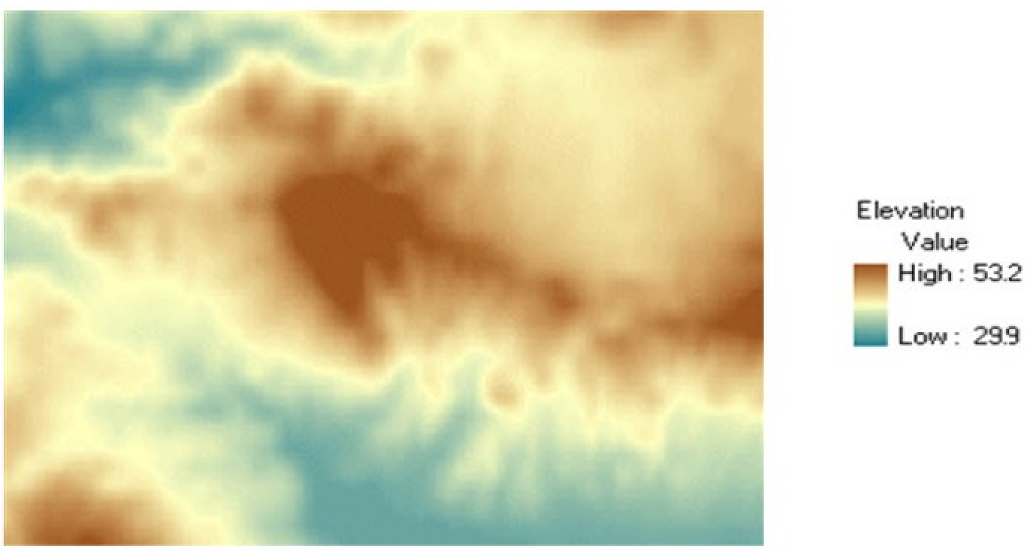
Raster datasets can represent non-geographic or geographic information. When rasters represent geographic information, they store the location and characteristics of each cell.

Rasters may be categorized as one of two types: image rasters and thematic rasters. Image rasters are typically produced by an optical or electronic device such as a camera or scanner. Digital photographs or images are a type of raster dataset.



This digital aerial photograph is composed of a grid of cells. Each cell represents a specific location on the earth.

Thematic rasters represent geographic features or phenomena with either discrete or continuous data. Digital elevation models (DEM) are a common type of thematic raster dataset.



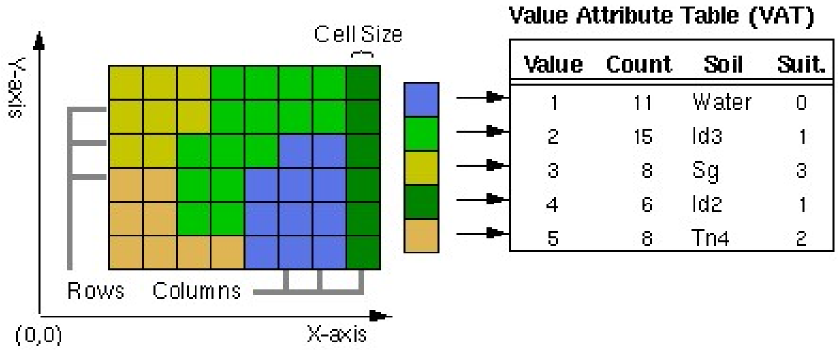
Each of the cells in a DEM raster represents the elevation of that point on the earth's surface.

Together, imagery and DEMs can provide a wealth of information about the topography of a region.

Raster Values

Raster datasets represent data using either integers (whole numbers) or floating points (numbers with decimals). Integer rasters are used to represent discrete data, such as landuse, and floating point rasters are used to represent continuous data, such as like elevation or slope.

A discrete raster dataset contains cells whose values are integers, often code numbers for a particular category. Cells can have the same value in a discrete raster dataset. Integer raster datasets have a value attribute table (VAT) that stores the cell values and their associated attributes.



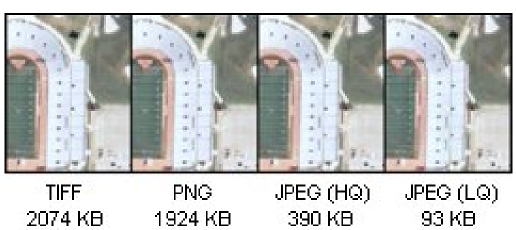
A VAT has one record for each unique value in the grid.

Unlike discrete raster datasets, continuous raster datasets (floating point raster datasets) do not have an attribute table because each cell in a continuous raster dataset can have a different floating point value. Cells in this type of raster dataset do not fall neatly into discrete categories.

Comparing Raster Formats

There are literally dozens of different formats in which to store raster data. Discussing each of the various proprietary raster file formats is well beyond the scope of this course. Rather, this course covers a few examples of some of the more common formats, and highlights the advantages and disadvantages of each.

The following graphic shows a zoomed-in portion of an aerial photo image in each of the common raster formats covered in this course. Image quality and file size will be compared for each of these formats.



This string of raster images shows the difference in image quality and file size for each of the raster formats covered in this course.

TIFF format

TIFF (tagged image file format) files are one of the preferred raster file types for geospatial data because they are suitable for representing multi-color images, but can also support black and white and gray-scale images. The TIFF format is based on lossless compression so it can be stored in a compressed or decompressed format without losing any of its original data. Another advantage of the TIFF format is its ability to store reference tags in the file that establishes a spatial reference for the TIFF file; such a tagged file is generally referred to as a "GeoTIFF".

PNG format

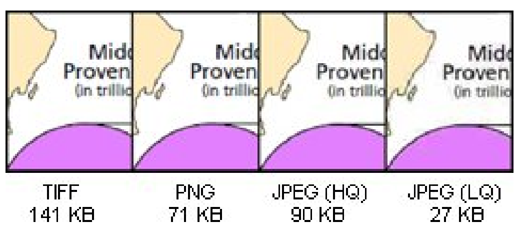
Another popular image format is the PNG (portable network graphic) format, which is sometimes thought of as the successor to the GIF (graphics interchange format). This format is excellent for simple raster images (such as the Mideast image below), and compresses well. The PNG is commonly found on the Internet, where it is quickly replacing the GIF as the standard image format. It is also used for exporting maps and maintaining raster attributes in geodatabases. Its increased support of transparency, lossless data compression, and increased color depth all make this an ideal format for web-based applications. It is important to note, however, PNG files do not support the CMYK color space, so they are rarely used in professional printing. Note, however, that the file sizes demonstrate that the PNG's lossless compression technique is minimally effective, as it produces a file only 7% smaller than the uncompressed TIFF.

JPEG format

JPEG (joint photographic experts group) files are commonly used on the Internet and produced by default on most digital cameras. JPEG format uses "lossy" compression to create as small a file size as possible while maintaining the overall quality of the image; that is, the image will look much like a raw, uncompressed image, but will not be a precise bit-by-bit copy. This works well for photographs, where minute changes are, for all practical purposes, imperceptible to the human eye. However, for applications that require precise data values, the JPEG format is not the best choice.

Between the four images above, the only noticeable difference in quality appears in the low-quality JPEG file. The low-quality JPEG brings the file size down considerably, but at a high cost to quality. The high-quality JPEG compression works quite well, however, as the file size is drastically reduced while the quality of the image remains basically unchanged.

Compare the following images of a map of the Middle East, derived from vector data.



This string of raster images was derived from vector data. It offers another chance to compare image quality and file size for each of the raster formats covered in this course.

The first thing you will notice is that all four images appear to be a little "blocky". This is simply because the image was converted from vector to raster format on the basis of how the map would look in its entirety. As you zoom in, you will begin to see the blockiness. If higher quality is desired for a zoomed-in portion of the image, the vector-to-raster conversion would be done with a higher PPI (pixels per inch) value to provide smoother features in the image.

GIS raster formats

Some GIS-specific raster formats are MrSID (multiresolution seamless image database), a proprietary format developed by LizardTech; ASCII-Grid format, which stores a raster dataset in a human-readable text-file format (suitable for sending via e-mail or USENET); and of course, Esri's own proprietary Esri GRID format.

ERDAS IMAGINE uses IMG files to store raster data, and is supported as one of the standard ArcGIS files. The contents of an IMG file are not fixed. Because of the open nature of the file format, other developers may create and add new types of items to the file.

Another common format is ECW (enhanced compression wavelet), which is a proprietary wavelet compression image format optimized for aerial and satellite imagery. This lossy compression format efficiently compresses large images with fine alternating contrast.

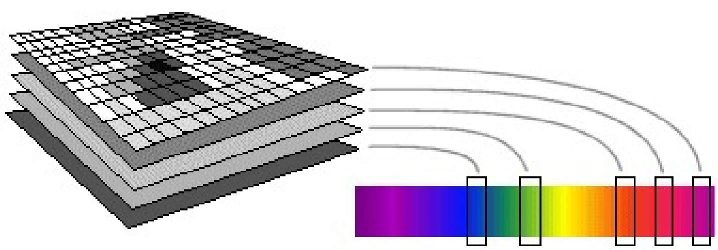
The raster file format you choose to use will depend on desired quality versus desired file size or available disk space, as well as the intended use of the file, whether it be for a website (PNG or JPEG) or to transfer high-quality geospatial data (TIFF).

Raster bands

A raster dataset can be either simple (one layer) or composite (a collection of multiple layers). These raster layers are referred to as bands.

Thematic rasters and panchromatic images are examples of simple raster datasets. Many types of imagery such as multispectral satellite imagery, which you will learn about in the next module, are composite rasters.

When there are multiple bands, every cell location has more than one value associated with it.



A raster dataset can have multiple bands. For example, each band in a multispectral image represents different parts of the electromagnetic spectrum. Every cell in the image has a data value for each band.

Cell values in single-band rasters are usually displayed using a gray-scale or a color map. In gray-scale, shades range between black and white. In a color map, cell values are arbitrarily matched to particular colors.

Some rasters also have an alpha band, which acts as a transparency mask, providing a transparency value for each pixel. ArcGIS supports alpha bands for the JPEG 2000 and PNG file formats.

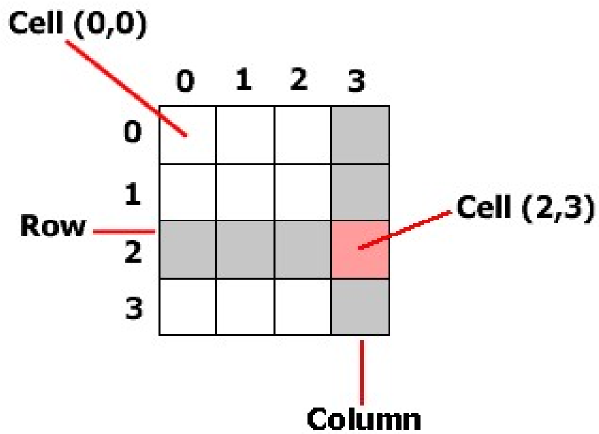
Spatial Reference for Rasters

A dataset's spatial reference describes where features are located in the real world. The spatial reference includes a coordinate system for x-, y-, and z-values as well as resolution and tolerance values for x-, y-, z-, and m- values.

Coordinate systems

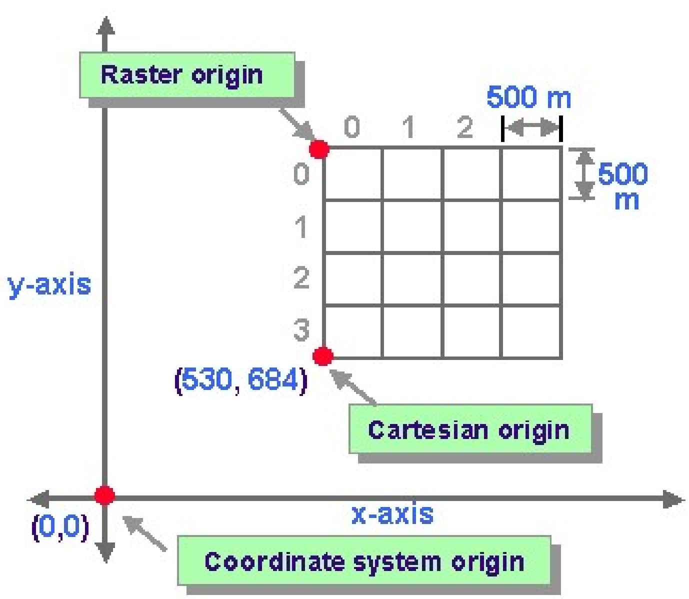
All geospatial raster datasets are in some coordinate space. This coordinate space may be a real-world coordinate system such as latitude and longitude, or one based on the raster's cells.

In the raster, cells are referenced by their row and column position. The rows and columns are numbered from the top left corner of the raster, starting with zero. The cell in the top left corner has row and column coordinates of 0,0. This cell is known as the raster's cell origin.



Every cell in a raster can be uniquely identified by its row and column position. For example, the red cell is at row 2 and column 3.

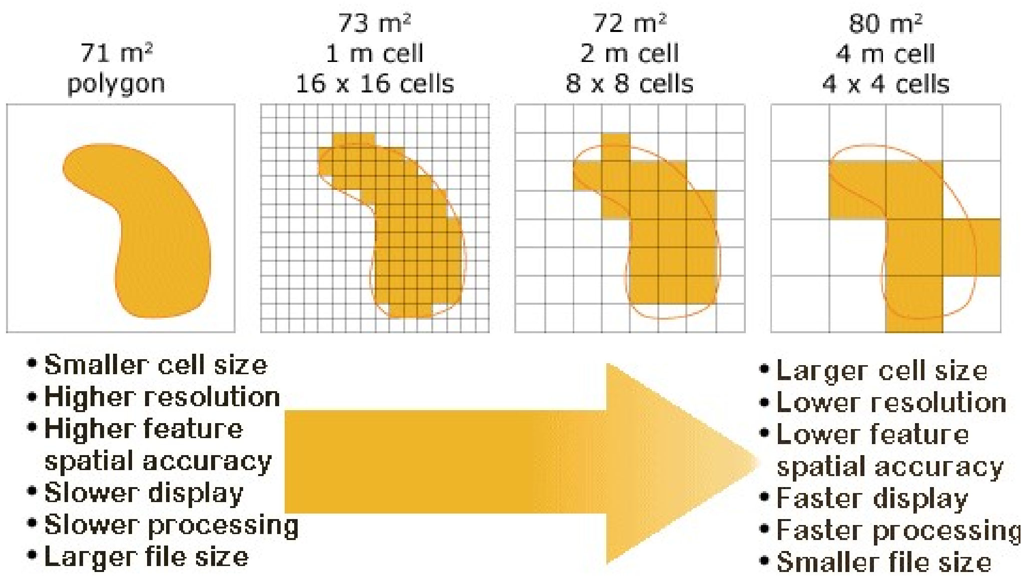
A raster that represents geographic information uses a Cartesian coordinate system to reference cells to a location on the earth. The referenced raster is always oriented so that the cell boundaries are parallel to the x- and y-axis of the Cartesian coordinate system. The bottom left corner of a raster serves as the starting point for referencing the raster with the real-world coordinates.



A raster that represents geographic information has a second origin for the Cartesian coordinate system, the bottom left corner.

Spatial Resolution, or Cell Size

Spatial resolution refers to the area of the real world represented by one cell in the raster. A high-resolution dataset will have cells that represent relatively small areas in the real world, thereby providing more detail. A low-resolution dataset will have cells that represent a large area and provide a "summary" of the area in question.



The level of detail (of features/phenomena) represented by a raster is often dependent on the cell size, or spatial resolution, of the raster. The cell must be small enough to capture the required detail but large enough so computer storage and analysis can be performed efficiently.

Tolerance

A spatial reference also includes tolerance values that reflect the accuracy of the coordinate data. The tolerance value is the minimum distance between coordinates. If one coordinate is within the tolerance value of another, they are interpreted as being at the same location. The default tolerance is set to 0.001 meter or its equivalent in map units. This is 10 times the default resolution value and is recommended in most cases.

Vector data

The vector format is an excellent way to store information about discrete geographical locations, features, or regions. While rasters often represent data that is continuous or thematic (e.g., images, temperature, precipitation), vector data is used more often to represent discrete objects such as trees, streets, or buildings.

In geodatabases, a vector object is known as a feature, and related features are organized into groups called feature classes. Unlike rasters, where information is stored within each cell of the raster, the attributes of vector data are associated with features.

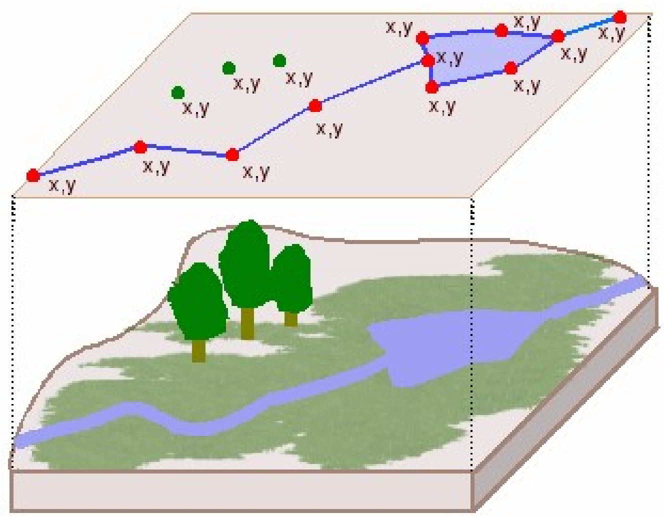
ArcMap displays what are called layers. A layer references the information contained in a feature class or raster, rather than actually storing it. Layers can be saved, but do not actually contain the information in the feature class or raster. Conversely, feature classes and rasters do not store symbology directly. Therefore, if you want to share information with other ArcGIS users and have it display in a specific way, you must send both the physical information (feature classes, shapefiles, rasters) and the symbology information (layer files). If you do not include layer files, the data will display with default ArcMapsymbology. If you do not include the physical information, there will be nothing to display in ArcMap.

Vector data model

The vector data model is based on the assumption that the earth's surface is composed of discrete objects such as trees, rivers, and lakes. Objects are represented as point, line, and polygon features with well-defined boundaries. Feature boundaries are defined by x,y coordinate pairs, which reference a location in the real world.

- Points are defined by a single x,y coordinate pair.
- Lines are defined by two or more x,y coordinate pairs.
- Polygons are defined by lines that close to form the polygon boundaries.

In the vector data model, every feature is assigned a unique numerical identifier, which is stored with the feature record in an attribute table.



The vector data model represents real-world features as points, lines, and polygons whose boundaries are defined by x,y coordinate pairs.

Feature attributes

On a GIS map, there is more to a feature than its location and shape. There is all the information associated with that feature. For a road, this might include its name, speed limit, and whether it is one-way or two-way. For a city, this might include its population, demographic characteristics, number of schools, and average monthly temperatures.

Information associated with a feature in a GIS is called an attribute. For example, population can be an attribute of a city, country, continent, and other features. Feature attributes are stored in an attribute table. In an attribute table, each feature is a record (row) and each attribute is a field (column). The attributes for all the features in a layer are stored in the same attribute table.

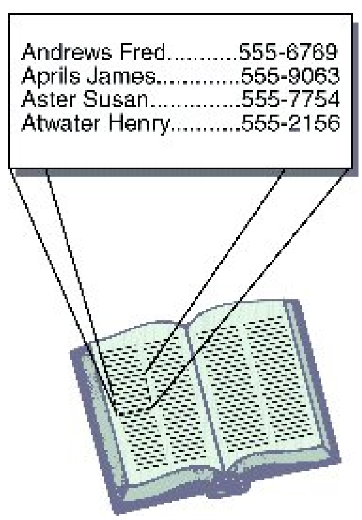
FID	Shape*	CITY_NAME	CNTRY_NAME	STATUS
0	Point	Drammen	Norway	Provincial capital
1	Point	Dundee	United Kingdom	Other
2	Point	Hunterston	United Kingdom	Other
3	Point	Ronne	Denmark	Provincial capital
4	Point	Petropavlovsk	Kazakhstan	Provincial capital
5	Point	Teesport	United Kingdom	Other
6	Point	Gdynia	Poland	Other
7	Point	Schwerin	Germany	Provincial capital
8	Point	Bremerhaven	Germany	Other
9	Point	Europoort	Netherlands	Other
10	Point	Dunkirk	France	Other

This attribute table for a layer of cities stores each feature's ID number, shape, name, the country in which it is located, and its status.

Each feature in a feature class—essentially, a group of related features—may have any given number of attributes. Each attribute value can be categorized as one of four levels of measurement: nominal (such as ID numbers or names), ordinal (rankings), interval (on a scale with no "natural" zero), or ratio (on a scale with a "natural" zero). The level of measurement defines how each attribute's values can be interpreted and used in analysis. Understanding these levels of measurement help you understand the features or phenomenon being modeled.

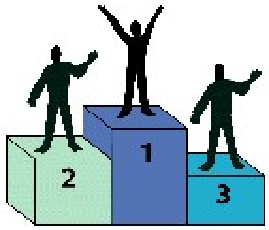
More about the four levels of measurement

The nominal level is the lowest level of measurement. Nominal data can only be distinguished qualitatively, such as soil type or telephone numbers. No order or ranking is implied. Using nominal data, features can only be compared as to whether they are equal or share membership in a particular group.

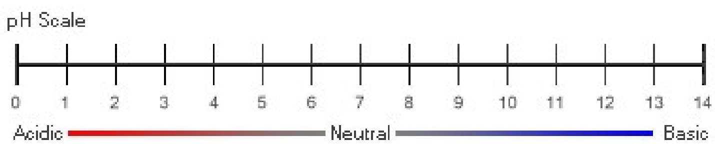


Names and telephone numbers are both nominal attributes, which can be used to identify a feature or to categorize it based on membership in a group.

Data at the **ordinal level** can be ranked into hierarchies, but the differences among measurements are not meaningful. Comparisons are implied, but not quantified, with such terms as "high, medium, low," "first, second, third," or "A, B, C." Examples of ordinal data include stream order or city hierarchies.

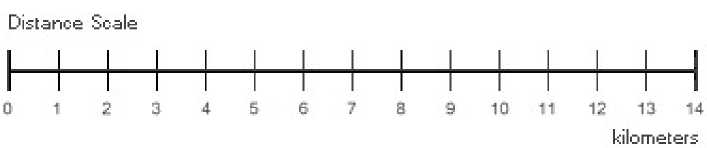


The interval level of measurement indicates the distance between the ranks of measured elements, but a starting point is arbitrarily assigned. Because there is no true zero point, relative comparisons can be made between the measurements, but ratio and proportion determinations are not as useful. Time of day, calendar years, and pH values are all examples of interval measurements. These are values on a linear calibrated scale, but they are not relative to a true zero point in time or space. The best example of interval measurement is Celsius temperature, in which 0 degrees Celsius is an arbitrary value, and 20 degrees Celsius is not twice as hot as 10 degrees Celsius.



Values on a pH scale are an example of interval data. The difference between values is meaningful, but a value of 0 on this scale is arbitrary and does not indicate a starting point.

Ratio measurements, the highest level of measurement, include an absolute starting point. The values from the ratio measurement system are derived relative to a fixed zero point on a linear scale. Mathematical operations can be used on these values with predictable and meaningful results. Examples of ratio measurements are age, distance, weight, and volume.



Distance values are examples of ratio attribute data. With ratio measurements, there is a zero starting point that is meaningful. Ratio measurements can be quantified and compared.

Vector formats

As with the raster data formats, there are a number of file formats for storing vector data. This module concentrates on the shapefile and the file geodatabase, which you learned about earlier. It will also briefly cover the KML (Keyhole Markup Language) format.

Shapefiles are universal exchange files (unlike the geodatabase). The term "shapefile" is somewhat misleading, as a shapefile actually comprises a group of separate files. ArcMap and ArcCatalog make these separate files invisible to the end user by displaying them as a single file with a SHP extension.

Like KML (KMZ) files, shapefiles are not stored in a geodatabase. You can, however, import the features in a shapefile into a feature class in a geodatabase.

Note: It is very important when moving or copying shapefiles that you use ArcCatalog to make certain the appropriate files are included. You will lose data if you do not do this.

Another format that has become popular is the Keyhole Markup Language (KML) file. KML files have either a KML file extension or a KMZ file extension (for compressed and zipped KML files). Each KML file is composed of a collection of graphic elements, images, and settings. KML is used to do the following:

- Symbolize and display GIS data as elements within Google Earth and Google Maps using symbols, color, images, and balloon-style information pop-ups.
- Provide access to attribute information about geographic features—for example, by presenting attribute information when you click a feature's symbol.
- Define the user's interaction with those features—for example, to control fly-to and camera location settings in Google Earth.

With ArcGIS, you may export any layer or map you create in ArcMap to KMZ format, which can then be loaded into ArcGIS and other applications, such as Google Earth.

There are a number of other vector formats that are occasionally used. Two of the more common formats are the AutoCAD Drawing Exchange Format (.DXF) and the United States Geological Survey (USGS) Digital Line Graph (.DLG) format. ArcGIS 9.3 supports the loading of the DXF format. The DLG format can be loaded with the Data Interoperability extension.

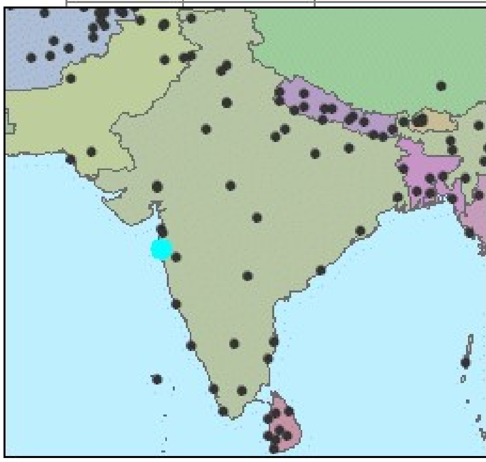
Data exploration

Exploring GIS data begins with the obvious panning, zooming, and toggling of layers that you have already employed in this lesson. In this topic, you will begin to explore some of the other simple data exploration techniques offered by ArcGIS, including querying attribute tables.

Attribute tables

A feature on a GIS map is linked to its record in the attribute table by a unique numerical identifier (ID). Every feature in a layer has an identifier. It is important to understand this one-to-one relationship between feature, identifier, and attribute record. Because features on the map are linked to their records in the table, you can click a feature on the map and see the attributes stored for it in the table. When you select a record in the table, the linked feature on the map is automatically selected as well. In the example below, each row in the table represents a city. Columns in the table represent different city attributes, such as its name, country, and capital status.

FID	Shape*	CITY_NAME	CNTRY_NAME	STATUS
694	Point	Bolu	Turkey	Provincial capital
1304	Point	Bombay	India	Provincial capital
2018	Point	Bombo	Uganda	Provincial capital
			Dominican Republic	Provincial capital
			Cote d'Ivory	Provincial capital
			Chad	Provincial capital
			Cote d'Ivory	Provincial capital
			Germany	Other
			Somalia	Provincial capital



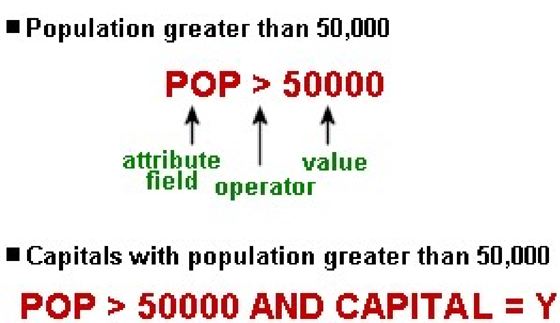
In this map of India, the city of Bombay is selected. Its record is also selected in the attribute table. The unique identifier for Bombay is stored in the FID field.

The Shape* field stores the feature geography. It is denoted with an asterisk because it contains elements that are not shown in the table.

Links between features and attributes make it possible to ask questions about the information stored in an attribute table and display the answer on the map. This linkage makes GIS maps much more informative than static maps. You will learn more about how to query attributes for information later in this module.

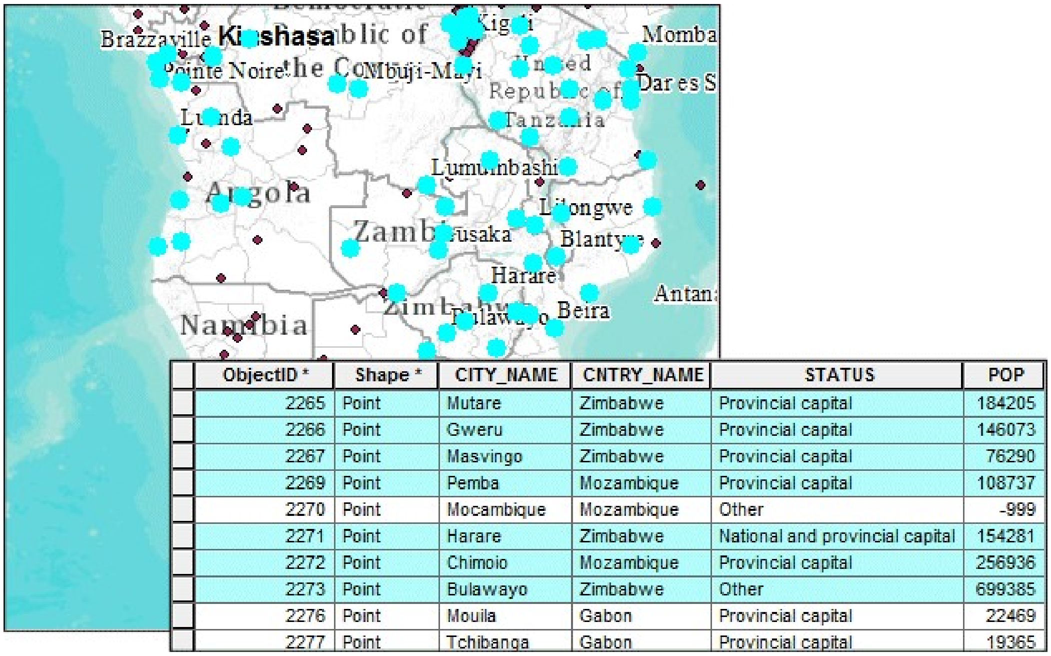
Using attribute queries

To find features that meet specific attribute criteria, you create a query expression. A query expression is a logical statement consisting of three parts: a field name (attribute), an operator, and an attribute value. Query expressions can be linked together to include multiple criteria. Expressions that contain multiple criteria are called compound expressions.



Query expressions can be simple, like the top example, or compound, like the bottom example. In this compound query expression, two attributes are being queried. The expression will select capital cities whose population is greater than 50,000.

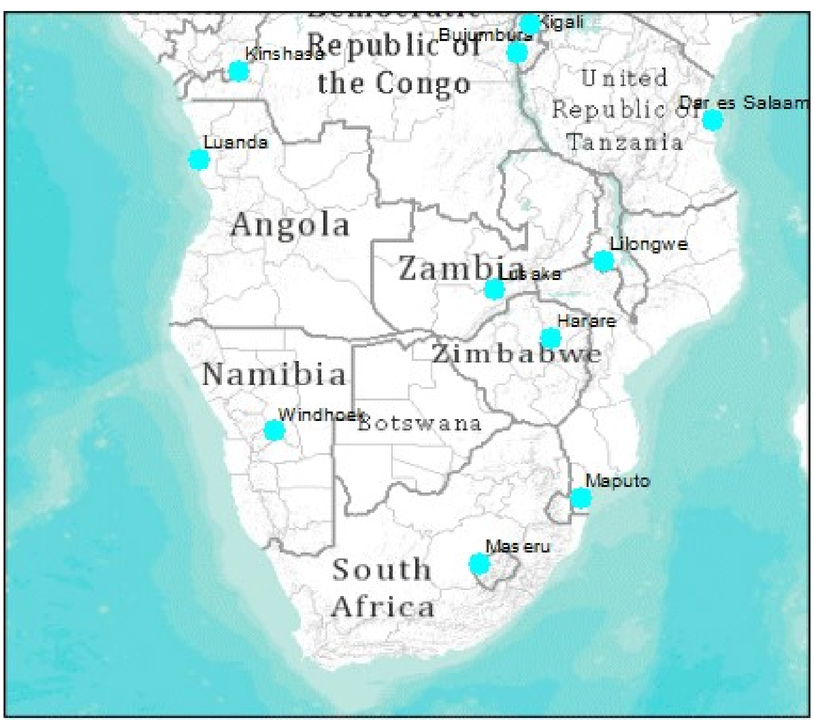
The result of the query is a selected set of the features in the layer. The features are selected on the map and their corresponding records are selected in the layer attribute table.



	ObjectID *	Shape *	CITY_NAME	CNTRY_NAME	STATUS	POP
	2265	Point	Mutare	Zimbabwe	Provincial capital	184205
	2266	Point	Gweru	Zimbabwe	Provincial capital	146073
	2267	Point	Masvingo	Zimbabwe	Provincial capital	76290
	2269	Point	Pemba	Mozambique	Provincial capital	108737
	2270	Point	Mocambique	Mozambique	Other	-999
	2271	Point	Harare	Zimbabwe	National and provincial capital	154281
	2272	Point	Chimoio	Mozambique	Provincial capital	256936
	2273	Point	Bulawayo	Zimbabwe	Other	699385
	2276	Point	Mouila	Gabon	Provincial capital	22469
	2277	Point	Tchibanga	Gabon	Provincial capital	19365

Cities with a population greater than 50,000 are selected on the map, and their records are selected in the layer attribute table. These features are the selected set resulting from the query expression POP > 50000.

In a special type of attribute query called a definition query, the query expression is a property of the layer and defines which features from the feature class will be included in the layer. A definition query is a way to eliminate features that are not of interest from the map display and also save processing time when the layer is queried or used in another GIS operation.



On this map, the Cities layer has a definition query that limits the layer features to only those cities with a population greater than 50,000 and that are national and provincial capitals. Notice that this map is much less cluttered than the one shown above.

Accessing more attributes

Not all feature attributes are stored in a layer attribute table. In many GIS databases, user-defined attributes (as opposed to software-generated attributes) are stored in separate, nonspatial tables. When you want to find features based on their attributes, sometimes you will need to associate a nonspatial table to the layer attribute table before you can perform the query.

You can associate a nonspatial table to a layer attribute table if they share a common field; that is, a field that stores the same data. The names of the common field do not have to be the same, but the field types (e.g., text, short integer) must be.

Layer attribute table

PROV	NAME	SQKM
AB	Alberta	661163.3
BC	British Columbia	946089.46
MB	Manitoba	609546.55
NT	Northwest Territories	1284344.72
NU	Nunavut	2073678.11
SK	Saskatchewan	641557.98
YT	Yukon Territory	483665.73
NB	New Brunswick	73005.64
NF	Newfoundland And Labrador	395344.4
PE	Prince Edward Island	5890.84
NS	Nova Scotia	55864.31
ON	Ontario	982628.82
QC	Quebec	1496272.47

Economic statistics table

PROVINCE	UNEMPLMT
AB	5.0
BC	7.9
MB	5.0
NT	0.0
NU	0.0
SK	6.1
YT	0.0
NB	9.9
NF	15.5
PE	12.0
NS	9.2
ON	7.1
QC	8.6

The layer attribute table and the economic statistics table both contain a text field that stores two-letter abbreviations for each Canadian province. The tables can be associated because they have this field in common.

ArcMap provides two methods for associating nonspatial tables to layer attribute tables: joins and relates. Which method you use depends on how the individual records in each table are related to one another. Table joins are designed for one-to-one or many-to-one relationships.

In a one-to-one relationship, each feature has one related record in the other attribute table. In a many-to-one relationship, multiple features may have the same related record. The Canadian provinces tables above are an example of a one-to-one relationship, because each province has one unemployment record. An example of a many-to-one relationship would be a layer attribute table containing records for several job training offices within each province, related to the same economic statistics table.

When tables are joined, the fields from one table are appended to the other table. Table joins are virtual —the tables appear connected in ArcMap, but the fields from the nonspatial table have not actually been added to the layer attribute table. If you want the data from both tables permanently combined into one, you can export the joined table to a new table or feature class.

Table joins are not permanent; you can remove a table join whenever you want.

1.7 SELF-ASSESSMENT TEST

What is datum? How do we calculate latitude and longitude of a place?

What is EMR? Discuss about the different stages of remote sensing.

Discuss bout the different approaches in change detection studies using remote sensed products.

What is GIS? Discuss about the different components of GIS.

1.8 SUMMARIES AND KEY POINTS

The following topics have been discussed in the present section -

- Cartography – nature, scope and development;
- Nature and types of Geoid; Concept of Datum with special reference to NAD, Everest and WGS-84; Principles of Spherical Trigonometry
- Principles and properties of UTM Projections
- Concept of Geoinformatics; Remote Sensing Platforms and Sensors;
- Nature of EMR, EMS, and interaction with atmosphere and surface materials;
- Different sensors and resolutions;
- Digital Image Processing – principles and approaches;
- Analytical Modelling in GIS, GPS-GIS integration

1.9 STUDY TIPS

Remote Sensing and Image Interpretation – Lillesand, Kiefer and Chipman seventh edition

Remote Sensing of the Environment: An Earth Resource Perspective - John R. Jensen Second Edition

Introduction to the Physics and Techniques of Remote Sensing- Charles Elachi and Jakob van Zyl

GROUP – GEO310T.2: RESEARCH METHODOLOGY

(Credit – 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

A. Introduction

Research may be defined as the systematic method of discovering new facts or of verifying old facts through sequence, inter-relationship, causal explanations.It is an art of scientific investigation. There are hundreds of scientific fields and disciplines, ranging from the Physical Sciences, to the Life Sciences, to the Social Sciences. The fundamental governing principal they all share in common, that distinguishes a science as a science is what's referred to as the scientific method.The method is expressed in the scientific research that constantly adds new knowledge and discoveries to the various realms and endeavours of the sciences.Research methodology is a systematic and organised way to solve the research problem.

B. Learning Objectives

After reading this SLM, you will be able to

- Bear an idea of research – what it means, why a research is carried out or you may say, the significance of carrying out a research
- It will also help you to know the different types of research.
- This SLM will help you to gain the concepts of research methodologies.
- You will be able to know different types of sampling procedures and this knowledge will guide you to choose the suitable sampling method to meet your research objectives.
- Finally, it will help you to explain the differences between reference and bibliography, and abstract, summary and synopsis.

C. Assessment of Prior Knowledge

Before starting the class, the learners may be asked the following questions:

- ✓ What do you mean by research?
- ✓ What are the objectives of carrying out a research?

D. Learning Activities

The learning process will involve Personal Contact Programmes, discussion, debate and interaction among students themselves, and students and teacher. During the Personal Contact Programmes, students may be assigned to prepare assignments on the issues of approaches of research, bibliography, referencing styles, hypothesis building etc. They can be asked to prepare charts on the differences between inductive and deductive mode of research, objective and subjective research, bibliography and references etc.

E. Feedback of Learning Activities

Once the learning process is completed, internal assessments will be conducted. On the basis of evaluation reports of the internal assessments, some areas of the syllabus will be refocused depending on students' requirements.

2.1 Approaches to Research

Research approaches are plans and the procedures for research that span the steps from broad assumptions to detailed methods of data collection, analysis, and interpretation. This plan involves several decisions, and they need not be taken in the order in which they make sense to me and the order of their presentation here. The overall decision involves which approach should be used to study a topic. Informing this decision should be the philosophical assumptions the researcher brings to the study; procedures of inquiry (called research designs); and specific research methods of data collection, analysis, and interpretation. The selection of a research approach is also based on the nature of the research problem or issue being addressed, the researchers' personal experiences, and the audiences for the study. Thus, in this book, research approaches, research designs, and research methods are three key terms that represent a perspective about research that presents information in a successive way from broad constructions of research to the narrow procedures of methods.

Inductive and Deductive Approaches

Inductive Approach (Inductive Reasoning)

Inductive approach, also known in inductive reasoning, starts with the observations and theories are proposed towards the end of the research process as a result of observation. Inductive research "involves the search for pattern from observation and the development of explanations – theories – for those patterns through series of hypotheses". No theories or hypotheses would apply in inductive studies at the beginning of the research and the researcher is free in terms of altering the direction for the study after the research process had commenced.

It is important to stress that inductive approach does not imply disregarding theories when formulating research questions and objectives. This approach aims to generate meanings from the data set collected in order to identify patterns and relationships to build a theory; however, inductive approach does not prevent the researcher from using existing theory to formulate the research question to be explored. Inductive reasoning is based on learning from experience. Patterns, resemblances and regularities in experience (premises) are observed in order to reach conclusions (or to generate theory).

- ❖ Inductive reasoning works from specific observations to broader generalizations and theories.
- ❖ Informally, we sometimes call this a "bottom up" approach.
- ❖ Conclusion is likely based on premises.
- ❖ It involves a degree of uncertainty

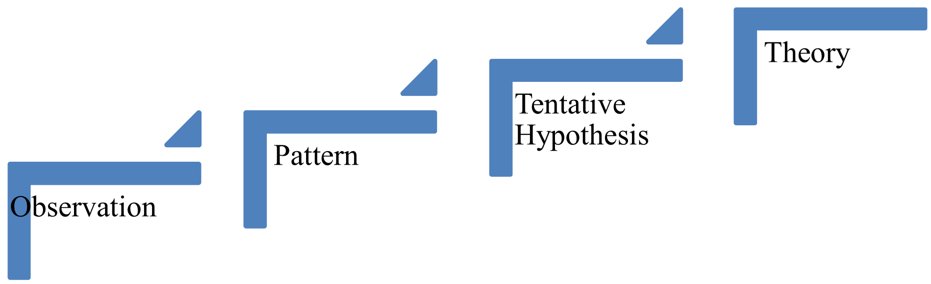


Figure: Inductive Approach

Inductive Approach for Analysing Qualitative Evaluation Data

According to Thomas (2006), the following procedures are used for the inductive analysis of qualitative data:

1. Preparation of raw data files (data cleaning): Format the raw data files in a common format (e.g., font size, margins, questions or interviewer comments highlighted) if required. Print and/or make a backup of each raw data file (e.g., each interview).

2. Close reading of text: Once text has been prepared, the raw text is read in detail until the evaluator is familiar with its content and gains an understanding of the themes and events covered in the text.

3. Creation of categories: The evaluator identifies and defines categories or themes. The upper-level or more general categories are likely to be derived from the evaluation aims. The lower-level or specific categories will be derived from multiple readings of the raw data, sometimes referred to as in vivo coding. In inductive coding, categories are commonly created from actual phrases or meanings in specific text segments. Several procedures for creating categories may be used. When using a word processor, marked text segments can be copied into the emerging categories. Specialist qualitative analysis software can be used to speed up the coding process when there are large amounts of text data (cf. Durkin, 1997).

4. Overlapping coding and uncoded text: Among the commonly assumed rules that underlie qualitative coding, two are different from the rules typically used in quantitative coding:

   (a) one segment of text may be coded into more than one category, and
   (b) a considerable amount of the text (e.g.,50% or more) may not be assigned to any category, because much of the text may not be relevant to the evaluation objectives.

5. Continuing revision and refinement of category system: Within each category, search for sub-topics, including contradictory points of view and new insights. Select appropriate quotations that convey the core theme or essence of a category. The categories may be combined or linked under a super ordinate category when the meanings are similar.

Example of the General Inductive Approach

In a qualitative evaluation of consumers' reports on their experiences with primary health care services in New Zealand, researchers collected consumers' accounts of their experiences using face-to-face, telephone, and focus group interviews (Kerse et al., 2004). A primary objective of the evaluation was to identify attributes of health care and encounters with health care providers that are considered important by consumers. Part of the interview schedule included the following items and questions:

• Tell me about a visit to your family doctor or other health service that went well for you.

• What made it go so well?

• Tell me about a visit that didn't go well.

• How could it have been improved?

The general inductive approach was used to analyze the qualitative data to identify themes in the text data that were related to the evaluation objectives. Once the data files were cleaned and put into a common format, the analysis commenced with a close reading of the text, which was carried out by two members of the evaluation team. During the analysis, specific themes were developed, which in the view of the investigators captured core messages reported by participants. An early example of a specific theme that emerged was labelled listening skills. Many of the respondents mentioned listening as being important during their encounters with primary health care providers. The following quotations are from text that was coded into this category.

<u>Examples of effective and ineffective listening are evident in these quotations</u>:

• "The way the doctor sits there and you explain, he listens to you."
• "The other doctor I had wouldn't listen. On one occasion when I was in a lot of pain he told me it was
all in my head."
• "This doctor listens and she is prepared to spend time."
• "He sorted all these out—I felt listened to, it was good; I felt happy."
• "Not in a hurry, ready to listen."

Other specific themes developed from the text included effective communication (e.g., "He is a good communicator and it is nice that he speaks the same language," "The doctor was using terms that I didn't understand"), providing clear information ("Ask any question and they answer what you want to know so you feel quite satisfied"), and giving advice on treatment effects ("They'll explain it and what it's likely to do and any side effects that could hurt"). These specific themes were then grouped into a broader category labelled communication.

The summary description for the communication category was "Good communication was seen to be essential to successful primary healthcare visits. Good communication by the health-care provider involved: good listening skills, using language that the consumer can understand, providing clear information and giving advice on treatment effects."

Four major themes were developed from the specific themes identified and described in the findings relating to consumer experiences. These themes constituted the primary framework for the findings relating to consumer experiences:

• communication (listening, information provided),
• relationship with service provider (equality, respect, approachability),
• availability of service provider, and
• professional expertise.

The model, consisting of the four major categories mentioned above, was used to highlight to stakeholders the main dimensions of quality in primary health care services from consumers' perspectives.

There are many good examples of inductive research, but we'll look at just a few here. One fascinating recent study in which the researchers took an inductive approach was Katherine Allen, Christine Kaestle, and Abbie Goldberg's study (2011)Allen, K. R., Kaestle, C. E., & Goldberg, A. E. (2011). More than just a punctuation mark: How boys and young men learn about menstruation. *Journal of Family Issues, 32*, 129–156. of how boys and young men learn about menstruation. To understand this process, Allen and her colleagues analyzed the written narratives of 23 young men in which the men described how they learned about menstruation, what they thought of it when they first learned about it, and what they think of it now. By looking for patterns across all 23 men's narratives, the researchers were able to develop a general theory of how boys and young men learn about this aspect of girls' and women's biology. They conclude that sisters play an important role in boys' early understanding of menstruation, that menstruation makes boys feel somewhat separated from girls, and that as they enter young adulthood and form romantic relationships, young men develop more mature attitudes about menstruation.

Deductive Approach

A deductive approach is concerned with "developing a hypothesis (or hypotheses) based on existing theory, and then designing a research strategy to test the hypothesis".
It has been stated that "deductive means reasoning from the particular to the general. If a causal relationship or link seems to be implied by a particular theory or case example, it might be true in many cases. A deductive design might test to see if this relationship or link did obtain on more general circumstances".

Deductive approach can be explained by the means of hypotheses, which can be derived from the propositions of the theory. In other words, deductive approach is concerned with deducting conclusions from premises or propositions.

Deduction begins with an expected pattern "that is tested against observations, whereas induction begins with observations and seeks to find a pattern within them"

- ❖ Deductive reasoning works from the more general to the more specific.
- ❖ Sometimes this is informally called a "top-down" approach.
- ❖ Conclusion follows logically from premises (available facts).

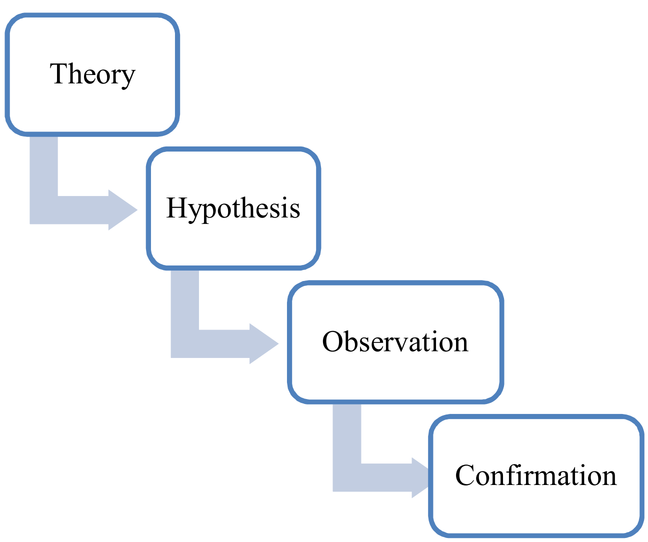


Figure: Deduvtive Approach

Advantages of Deductive Approach

Deductive approach offers the following advantages:

- Possibility to explain causal relationships between concepts and variables
- Possibility to measure concepts quantitatively
- Possibility to generalize research findings to a certain extent

Deductive research approach explores a known theory or phenomenon and tests if that theory is valid in given circumstances. It has been noted that "the deductive approach follows the path of logic most closely. The reasoning starts with a theory and leads to a new hypothesis. This hypothesis is put to the test by confronting it with observations that either lead to a confirmation or a rejection of the hypothesis".

Moreover, deductive reasoning can be explained as "reasoning from the general to the particular", whereas inductive reasoning is the opposite. In other words, deductive approach involves formulation of hypotheses and their subjection to testing during the research process, while inductive studies do not deal with hypotheses in any ways.

Generally, studies using deductive approach follow the following stages:

1) Deducing hypothesis from theory.

2) Formulating hypothesis in operational terms and proposing relationships between two specific variables

3) Testing hypothesis with the application of relevant method(s). These are quantitative methods such as regression and correlation analysis, mean, mode and median and others.

4) Examining the outcome of the test, and thus confirming or rejecting the theory. When analysing the outcome of tests, it is important to compare research findings with the literature review findings.

5) Modifying theory in instances when hypothesis is not confirmed.

Table 1: Differences between Inductive and Deductive Approach

Inductive	Deductive
Induction is usually described as moving from the specific to the general	Deduction begins with the general and ends with the specific.
Observations tend to be used for Inductive Arguments.	Arguments based on laws, rules and accepted principles are generally used for Deductive Reasoning.

Table 2: Nature of the Data and Research Approach

Approach	Qualitative	Quantitative
Inductive	Grounded theory	Exploratory data analysis
Deductive	Qualitative comparative analysis	Structural equating model

The main difference between inductive and deductive approaches to research is that whilst a deductive approach is aimed and testing theory, an inductive approach is concerned with the generation of new theory emerging from the data.

A deductive approach usually begins with a hypothesis, whilst an inductive approach will usually use research questions to narrow the scope of the study.

For deductive approaches the emphasis is generally on causality, whilst for inductive approaches the aim is usually focused on exploring new phenomena or looking at previously researched phenomena from a different perspective.

Inductive approaches are generally associated with qualitative research, whilst deductive approaches are more commonly associated with quantitative research. However, there are no set rules and some qualitative studies may have a deductive orientation.

One specific inductive approach that is frequently referred to in research literature is grounded theory, pioneered by Glaser and Strauss.

This approach necessitates the researcher beginning with a completely open mind without any preconceived ideas of what will be found. The aim is to generate a new theory based on the data.

Once the data analysis has been completed the researcher must examine existing theories in order to position their new theory within the discipline.

Grounded theory is not an approach to be used lightly. It requires extensive and repeated sifting through the data and analysing and re-analysing multiple times in order to identify new theory. It is an approach best suited to research projects where there the phenomena to be investigated has not been previously explored.

The most important point to bear in mind when considering whether to use an inductive or deductive approach is firstly the purpose of your research; and secondly the methods that are best suited to either test a hypothesis, explore a new or emerging area within the discipline, or to answer specific research questions.

Importance of Inductive and Deductive Approaches in Geographical Researches

Likewise researches in other social sciences, geographical research also walk with inductive and deductive approaches of reasoning. In geographical research earlier studies were accompanied with deductive method, where conclusions were drawn out from the premises. There were no serious attempts, by any geographer, made in past to test the validity of the existing premises. This was because of the reliability of the geographers on metaphysical world views (teleological approaches. It was believed that the changes in the phenomena were caused by law of nature.

However there were geographers who put emphasis on empirical observation and studied geography inductively. Modern geographical time period starts with two great German geographers viz. Alexander von Humboldt and Carl Ritter. Both were contemporary to each other but each had different approach to explore the geography. Humboldt was a believer of inductive approach and considered scientific observation as the sole tool for exploring the nature. He made several expeditions and field works to many places of South America, North America, Europe and Asia and came with many theories in the field of climatology, geomorphology, zoology, botany, geology etc. His book *Kosmos* is a comprehensive account of his journeys and expeditions. On the contrary, Carl Ritter, through fieldworks and empirical observation, put emphasis on deducing new conclusions from fundamentals assumptions or from already established truths. Thus his method was followed by deductive approach.

In geographical studies, there has always been methodological problem of dualism and dichotomy. Majority of the researches in the fields of geomorphology, climatology, oceanography and earth sciences are done using inductive approach, while specialized branches of geography rely on deductive methods. However both the approaches are used in researches irrespective of their fields in geography. It depends on the nature and objectives of the research.

While conducting research it is essential to determine whether to use and inductive or deductive approach. For this the researcher, geographer in this case, has to consider the purpose of the study first, and secondly the methods which are best suited for either testing hypothesis or to find answer of specific research questions.

2.2 Objective and Subjective Research

Research in social science requires the collection of data in order to understand a phenomenon. This can be done in a number of ways, and will depend on the state of existing knowledge of the topic area. The researcher can apply both objective and subjective methods.

Objective Research

Objective research tends to be modelled on the methods of the natural sciences such as experiments or large scale surveys. Objective research seeks to establish law-like generalisations which can be applied to the same phenomenon in different contexts. This perspective, which privileges objectivity, is called positivism and is based on data that can be subject to statistical analysis and generalisation. Positivist researchers use quantitative methodologies, which are based on measurement and numbers, to collect and analyse data. An objective research generally starts with a hypothesis and at the end, it test the hypothesis and explores the significance of the research.

Objective research tends to be modelled on the methods of the natural sciences such as experiments or large scale surveys. Objective research seeks to establish law-like generalisations which can be applied to the same phenomenon in different contexts. This perspective, which privileges objectivity, is called **positivism** and is based on data that can be subject to statistical analysis and generalisation. Positivist researchers use quantitative methodologies, which are based on measurement and numbers, to collect and analyse data. Interpretivists are more concerned with language and other forms of qualitative data, which are based on words or images. Having said that, researchers using objectivist and positivist assumptions sometimes use qualitative data while interpretivists sometimes use quantitative data. (Quantitative and qualitative methodologies will be discussed in more detail in the final part of this course.) The key is to understand the perspective you intend to adopt and realise the limitations and opportunities it offers. Table 1 compares and contrasts the perspectives of positivism and interpretivism.

Discussions involving the antagonism between objectivity and subjectivity often suffer from the fact that objectivity means different things to different people, in statistics and elsewhere (much of the following discussion will focus on the term “objectivity”; subjectivity is often considered as the opposite of objectivity and as such implicitly defined). Ambiguity in these terms is often ignored. We believe that such discussions can become clearer by referring to the meanings that are relevant in any specific situation instead of using the ambiguous terms “objectivity” and “subjectivity” without further explanation. The core of the current use of the term “objectivity” is the idea of impersonality of scientific statements and procedures.

According to Daston and Galison (2007), the term has only been used in this way in science from the mid-nineteenth century; before then, “objective” and “subjective” were used with meanings almost opposite from the current ones and did not play a strong role in discussions about science. The idea of independence of the individual subject can be applied in various ways. Megill (1994) listed four basic senses of objectivity: “absolute objectivity” in the sense of “representing the things as they really are” (independently of an observer), “disciplinary objectivity” referring to a consensus among experts within a discipline and highlighting the role of communication and negotiation, “procedural objectivity” in the sense of following rules that are independent of the individual researcher, and “dialectical objectivity.”

The latter somewhat surprisingly involves subjective contributions, because it refers to active human “objectification” required to make phenomena communicable and measurable so that they can then be treated in an objective way so that different subjects can understand them in the same way. Statistics for example relies on the construction of well delimited populations and categories within which averages and probabilities can be defined; see Desrosieres (2002).

Daston and Galison (2007) call the ideal of scientific images that attempt to capture reality in an unmanipulated way “mechanical objectivity” as opposed to “structural objectivity,” which emerged from the insight of scientists and philosophers such as Helmholtz and Poincare that observation of reality cannot exclude the observer and will never be as reliable and pure as “mechanical objectivists” would hope. Instead, “structural objectivity” refers to mathematical and logical structures. Porter (1996) lists the ideal of impartiality of observers as another sense of objectivity, and highlights the important role of quantitative and formal reasoning for concepts of objectivity because of their potential for removing ambiguities.

To us, the most problematic aspect of the term “objectivity” is that it incorporates normative and descriptive aspects, and that these are often not clearly delimited. For example, a statistical method that does not require the specification of any tuning parameters is objective in a descriptive sense (it does not require decisions by the individual scientist). Often this is presented as an advantage of the method without further discussion, implying objectivity as a norm, but depending on the specific situation the lack of flexibility caused by the impossibility of tuning may actually be a disadvantage (and indeed can lead to subjectivity at a different point in the analysis, when the analyst must make the decision of whether to use an auto-tuned approach in a setting where its inferences do not appear to make sense).

The frequentist interpretation of probability is objective in the sense that it locates probabilities in an objective world that exists independently of the observer, but the definition of these probabilities requires a subjective definition of a reference set. Although some proponents of frequentism consider its objectivity (in the sense of impersonality, conditional on the definition of the reference set) as a virtue, this property is ultimately only descriptive; it does not imply on its own that such probabilities indeed exist in the objective world, nor that they are a worthwhile target for scientific inquiry. The interpretation of objectivity as a scientific virtue is connected to what are seen to be the aims and values of science. Scientific realists hold that finding out the truth about the observerindependent reality is the major aim of science. This makes “absolute objectivity” as discussed above a core scientific ideal, as which it is still popular. But observer-independent reality is only accessible through human observations, and the realist ideal of objectivity has been branded as metaphysical, meaningless, and illusory by positivists, among which Porter (1996) counted Karl Pearson, and more contemporarily by empiricists such as van Fraassen (1980).

In the latter groups, objectivity is seen as a virtue as well, although for them it does not refer to observer-independent reality but rather to a standardized, disciplined, and impartial application of scientific methodology enabling academic consensus about observations. Reference to observations is an element that the empiricist, positivist, and realist ideas of objectivity have in common; Mayo and Spanos (2010) interpreted objectivity in a realist manner, and they saw checking theories against experience by means of what they call “error statistics” as central tool to ensure objectivity. In contrast, van Fraassen (1980) took observability and the ability of theory to account for observed facts as objective from an anti-realist perspective. His construal of observability depends on the context, theory, and means of observation and his concept of objectivity is conditional on these conditions of observation, assuming that at least acceptance of observations and observability given these conditions should not depend on the subject.

Daston and Galison (2007) portray the rise of "mechanical objectivity" as a scientific virtue in reaction to shortcomings of the earlier scientific ideal of "truth-to-nature," which refers to the idea that science should discover and present an underlying ideal and universal (Platonic) truth below the observed phenomena. The move towards mechanical objectivity, inspired by the development of photographic techniques, implied a shift of perspective; instead of producing pure and ideal "true" types the focus moved to capturing nature "as it is," with all irregularities and variations that had been suppressed by a science devoted to "truth-to-nature." Increasing insight in the shortcomings and the theory-dependence of supposedly objective observational techniques led to the virtue of "trained judgement" as a response to mechanical objectivity.

According to Daston and Galison (2007), the later virtues did not simply replace the older ones, but rather supplemented them, so that nowadays all three still exist in science. Daston and Galison do not discuss statistics, but the statistical idea of modeling error causing variation around a true parameter can be seen as an attempt to integrate the mechanically objective observation of variation with the ideal of truth-tonature. The idea from sampling theory of estimating a population quantity is more similar to the 9 photographic idea of estimating nature as it is. One could also connect subjective and objective Bayesian perspectives with the concept of trained judgement.

In Daston and Galison's framework, objectivity appears as one scientific virtue among others. Objectivity has also been criticized on the grounds that, as attractive as it may seem as an ideal, it is illusory. This criticism has to refer to a specific interpretation of objectivity, and a weaker interpretation of objectivity may still seem to critics to be a good thing: van Fraassen agrees with Kuhn (1962) and others that "absolute objectivity" is an illusion and that access to reality is dependent of the observer, but he still holds that objectivity conditional on a system of reference is a virtue. But there is even criticism of the idea that objectivity, possible or not, is desirable.

From a particular feminist point of view, MacKinnon (1987) wrote: "To look at the world objectively is to objectify it." Striving for objectivity itself is seen here as a specific and potentially harmful perspective, implying a denial of the specific conditions of an observer's point of view. A similar point was made by Feyerabend (1978). Maturana (1988) critically discussed the "explanatory path of objectivity-without-parenthesis" in which observers deny personal responsibility for their positions based on a supposedly privileged access to an objective reality; he accepted a more positive perspective-dependent use of the term called "objectivity-with-parenthesis."

Subjective Research

Subjective research generally refers to the subjective experiences of research participants and to the fact that the researcher's perspective is embedded within the research process, rather than seen as fully detached from it. Subjective research is generally referred to as phenomenological research. This is because it is concerned with the study of experiences from the perspective of an individual, and emphasises the importance of personal perspectives and interpretations. Subjective research is generally based on data derived from observations of events as they take place or from unstructured or semi-structured interviews. In unstructured interviews the questions emerged from the discussion between the interviewer and the interviewee. In semi-structured interviews the interviewer prepares an outline of the interview topics or general questions, adding more as needs emerged during the interview.

Subjective research is generally referred to as **phenomenological** research. This is because it is concerned with the study of experiences from the perspective of an individual, and emphasises the importance of personal perspectives and interpretations. Subjective research is generally based on data derived from observations of events as they take place or from unstructured or semi-structured interviews. In unstructured interviews the questions emerged from the discussion between the interviewer and the interviewee. In semi-structured interviews the interviewer prepares an outline of

the interview topics or general questions, adding more as needs emerged during the interview. Structured interviews include the full list of questions. Interviewers do not deviate from this list. Subjective research can also be based on examinations of documents. The researcher will attribute personal interpretations of the experiences and phenomena during the process of both collecting and analysing data. This approach is also referred to as **interpretivist** research. **Interpretivists** believe that in order to understand and explain specific management and HR situations, one needs to focus on the viewpoints, experiences, feelings and interpretations of the people involved in the specific situation.

Table 3: Differences between Objective and Subjective Research

Objective	Subjective
seeks to establish law-like generalisations	Explores facts and phenomena in a documented way
Generally starts with a hypothesis	May or may not start with hypothesis
Questionnaire is structured.	Questionnaire is unstructured.
Questions are close ended.	Questions are open-ended.
Applied mostly in natural sciences	Basically applied in social sciences
Belongs to positivist philosophy	Related to phenomenology and post-positivist approach

Key Differences between Objective and Subjective Research

The fundamental differences between objective and subjective are discussed in the given below points:

a) A neutral statement, which is completely true and real, unbiased and balanced, is an objective one. Subjective means something which does not show the clear picture or it is just a person's outlook or expression of opinion.

b) An objective statement is based on facts and observations. On the other hand, a subjective statement relies on assumptions, beliefs, opinions and influenced by emotions and personal feelings.

a) Objective information is provable, measurable and observable. In contrast, subjective information is relative to the subject, i.e. the person making it.

b) The objective statement can be checked and verified. Unlike subjective statement or a series of balanced opinions, so they can't be checked and verified.

c) When a piece of information is objective, it remains same, irrespective of the person reporting it. Conversely, a subjective statement differs from individual to individual.

d) An objective statement is appropriate for decision making, which is not in the case of a subjective statement.

e) You can find the objective statement in hard science, textbooks and encyclopaedias, but a subjective statement is used in blogs, biographies, and comments on social media.

Our attitude toward objectivity and subjectivity in science

The attitude taken in the present paper is based on Hennig (2010), which was in turn inspired by constructivist philosophy (Maturana, 1988, von Glasersfeld, 1995) and distinguishes personal reality, social reality, and observer-independent reality. According to this perspective, human inquiry starts from observations that are made by personal observers (personal reality). Through communication, people share observations and generate social realities that go beyond a personal point of view. These shared realities include for example measurement procedures that standardize observations, and mathematical models that connect observations to an abstract formal system that is meant to create a thought system cleaned from individually different point of views.

Nevertheless, human beings only have access to observer-independent reality through personal observations and how these are brought together in social reality. According to Hennig (2010), science aims at arriving at a view of reality that is stable and reliable and can be agreed freely by general observers and is therefore as observer-independent as possible. In this sense we see objectivity as a scientific ideal. But at the same time we acknowledge what gave rise to the criticism of objectivity: the existence of different individual perspectives and also of perspectives that differ between social systems, and therefore the ultimate inaccessibility of a reality that is truly independent of observers, is a basic human condition.

Objectivity can only be attributed by observers, and if observers disagree about what is objective, there is no privileged position from which this can be decided. Ideal objectivity can never be achieved. This does not imply, however, that scientific disputes can never be resolved by scientific means. Yes, there is an element of "politics" involved in the adjudication of scholarly disagreements, but, as we shall discuss, the norm of transparency and other norms associated with both objectivity and subjectivity can advance such discussions. In general no particular observer has a privileged position but this does not mean that all positions are equal. We recognize subjectivity not to throw up our hands and give up on the possibility of scientific consensus but as a first step to exploring and, ideally, reconciling, the multiple perspectives that are inevitable in nearly any human inquiry. Denying the existence of different legitimate subjective perspectives and of their potential to 10 contribute to scientific enquiry cannot make sense in the name of objectivity. Heterogeneous points of view cannot be dealt with by imposing authority.

Our attitude values the attempt to reach scientific agreement between different perspectives, but ideally such an agreement is reached by free exchange between the different points of view. In practice, however, agreement will not normally be universal, and in order to progress, science has to aim at a more restricted agreement between experts who have enough background knowledge to either make sure that the agreement about something new is in line with what was already established earlier, or to know that and how it requires a revision of existing knowledge. But the resulting agreement is still intended to be potentially open for everyone to join or to challenge. Therefore, in science there is always a tension between the ideal of general agreement and the reality of heterogeneous perspectives.

Furthermore our attitude to science is based on the idea that consensus is possible regarding stable and reliable statements about the observed reality (which may require elaborate measurement procedures), and that science aims at nontrivial knowledge in the sense that it makes statements about observable reality that can and should be checked and potentially falsified by observation. Although there is no objective access to observer-independent reality, we acknowledge that there is an almost universal human experience of a reality perceived as located outside the observer and as not controllable by the observer. This reality is a target of science, although it cannot be taken for granted that it is indeed independent of the observer. We are therefore "active scientific realists" in the sense of Chang (2012), who writes: "I take reality as whatever is not subject to one's will, and knowledge as an ability to act without being frustrated by resistance from reality. This perspective allows an optimistic rendition of the pessimistic induction, which celebrates the fact that we can be successful in science without even knowing the truth. The standard realist argument from success to truth is shown to be ill-defined and flawed."

This form of realism is not in contradiction to the criticism of realism by van Fraassen or the arguments against the desirability of certain forms of objectivity by constructivists or feminists as outlined above. Active scientific realism implies that finding out the truth about objective reality is not the ultimate aim of science, but that it rather supports human actions. This means that scientific methodology has to be assessed relative to the specific aims and actions connected to its use. Another irreducible subjective element in science, apart from multiple perspectives on reality, is therefore the aim of scientific inquiry, which cannot be standardized in an objective way.

A typical statistical instance of this is how much prediction accuracy in a restricted setting is valued compared with parsimony and interpretability. Because science aims at agreement, communication is central to science, as are transparency and techniques for supporting the clarity of communication. Among these techniques are formal and mathematical language, standardized measurement procedures, and scientific models. Objectivity as we see it is therefore a scientific ideal that can never fully be achieved. As much as science aims for objectivity, it has to acknowledge that it can only be built from a variety of subjective perspectives through communication.

Subjectivist Bayesianism

We call "subjectivist epistemic" the interpretation of probabilities as quantifications of strengths of belief of an individual, where probabilities can be interpreted as derived from, or implementable through, bets that are coherent in that no opponent can cause sure losses by setting up some combinations of bets. From this requirement of coherence, the usual probability axioms follow (O2c). Allowing conditional bets implies Bayes's theorem, and therefore, as far as inference concerns learning from observations about not (yet) observed hypotheses, Bayesian methodology is used for subjectivist epistemic probabilities, hence the term "subjectivist Bayesianism."

A major proponent of subjectivist Bayesianism was Bruno de Finetti (1974). De Finetti was not against objectivity in general. He viewed observed facts as objective, as well as mathematics and logic and certain formal conditions of random experiments such as the set of possible outcomes. But he viewed uncertainty as something subjective and he held that objective (frequentist) probabilities do not exist. He claimed that his subjectivist Bayesianism appropriately takes into account both the objective (see above) and subjective (opinions about unknown facts based on known evidence) components for probability evaluation. Given the degree of idealization required for frequentism as, this is certainly a legitimate position. In de Finetti's work the term "prior" refers to all probability assignments made before seeing the data, with no fundamental distinction between the "parameter prior" assigned to parameters in a model, and the form of the "sampling distribution" given a fixed parameter, in contrast to common Bayesian practice today, in which the term "prior" is used to refer only to the parameter prior (Gelman and Henning, 2015).

In the following discussion we shall use the term "priors" in de Finetti's general sense. Regarding the list of virtues in Section 4.2, de Finetti provided a clear definition of probability (O1a) based on principles that he sought to establish as generally acceptable (O2c). As opposed to objectivist Bayesians, subjectivist Bayesians do not attempt to enforce agreement regarding prior distributions, not even given the same evidence; still, de Finetti (1974) and other subjectivist Bayesians proposed rational principles for assigning prior probabilities. The difference between the objectivist and subjectivist Bayesian point of view is rooted in the general tension in science explained above; the subjectivist approach can be criticized for not supporting agreement enough— conclusions based on one prior may be seen as irrelevant for somebody who holds another one (O2c)—but can be defended for honestly acknowledging that prior information often does not come in ways that allow a unique formalization (S2b).

In any case it is vital that subjectivist Bayesians explain transparently how they arrive at their priors, so that other researchers can decide to what extent they can support the conclusions (O1c). Such transparency is desirable in any statistical approach but is particularly relevant for subjective Bayesian models which cannot be rejected within the subjectivist paradigm in case of disagreement with observations. 16 In de Finetti's conception, probability assessments, prior and posterior, can ultimately only concern observable events, because bets can only be evaluated if the experiment on which a bet is placed has an observable outcome, and so there is a clear connection to observables (O3a). However, priors in the subjectivist Bayesian conception are not open to falsification (O3b), because by definition they have to be fixed before observation. Adjusting the prior after having observed the data to be analyzed violates coherence (Gelman and Henning, 2015).

The Bayesian system as derived from axioms such as coherence (as well as those used by objectivist Bayesians; see Section 5.4) is designed to cover all aspects of learning from data, including model selection and rejection, but this requires that all potential later decisions are already incorporated in the prior, which itself is not interpreted as a testable statement about yet unknown observations. In particular this means that once a subjectivist Bayesian has assessed a setup as exchangeable a priori, he or she cannot drop this assumption later, whatever the data are (think of observing twenty zeroes, then twenty ones, then ten further zeroes in a binary experiment). This is a major problem, because subjectivist Bayesians use de Finetti's theorem to justify working with parameter priors and sampling models under the assumption of exchangeability, which is commonplace in Bayesian statistics.

Dawid (1982) discussed calibration (quality of match between predictive probabilities and the frequency of predicted events to happen) of subjectivist Bayesians inferences, and he suggests that badly calibrated Bayesians could do well to adjust their future priors if this is needed to improve calibration, even at the cost of violating coherence. Subjectivist Bayesianism scores well on the subjective virtues S1 and S2b. But it is a limitation that the prior distribution exclusively formalizes belief; context and aims of the analysis do not enter unless they have implications about belief. In practice, an exhaustive elicitation of beliefs is rarely feasible, and mathematical and computational convenience often plays a role in setting up subjective priors, despite de Finetti's having famously accused frequentists of "adhockeries for mathematical convenience." Furthermore, the assumption of exchangeability will hardly ever precisely match an individual's beliefs in any situation—even if there is no specific reason against exchangeability in a specific setup, the implicit commitment to stick to it whatever will be observed seems too strong—but some kind of exchangeability assumption is required by Bayesians for the same reason for which frequentists need to rely on independence assumptions: some internal replication in the model is needed to allow generalization or extrapolation to future observations.

Summarizing, we view much of de Finetti's criticism of frequentism as legitimate, and subjectivist Bayesianism comes with a commendable honesty about the impact of subjective decisions and allows for flexibility accommodating multiple perspectives. But checking and falsification of the prior is not built into the approach, and this can get in the way of agreement between observers. Furthermore, some problems of the frequentist approach criticized by de Finetti and his disciples stem from the unavoidable fact that useful mathematical models idealize and simplify personal and social perspectives on reality (see Hennig, 2010 and above), and the subjectivist Bayesian approach incurs such issues as well.

Objectivist Bayesianism

Given the way objectivity is often advertised as a key scientific virtue (often without specifying what exactly it means), it is not surprising that de Finetti's emphasis on subjectivity is not shared by all Bayesians, and that there have been many attempts to specify prior distributions in a more objective way. Currently the approach of E. T. Jaynes (2003) seems to be among the most popular. As with many of his predecessors such as Jeffreys and Carnap, Jaynes saw probability as a generalization of binary logic to uncertain propositions. Cox (1961) proved that given a certain list 17 of supposedly common-sense desiderata for a "plausibility" measurement, all such measurements are equivalent, after suitable scaling, to probability measures.

This theorem is the basis of Jaynes' objectivist Bayesianism, and the claim to objectivity comes from postulating that, given the same information, everybody should come to the same conclusions regarding plausibilities: prior and posterior probabilities (O2c), a statement with which subjectivist Bayesians disagree. In practice, this objectivist ideal seems to be hard to achieve, and Jaynes (2003) admits that setting up objective priors including all information is an unsolved problem. One may wonder whether his ideal is achievable at all. For example, in chapter 21, he gives a full Bayesian "solution" to the problem of dealing with and identifying outliers, which assumes that prior models have to be specified for both "good" and "bad" data (between which therefore there has to be a proper distinction), including parameter priors for both models, as well as a prior probability for any number of observations to be "bad."

It is hard to see, and no information about this is provided by Jaynes himself, how it can be possible to translate the unspecific information of knowing of some outliers in many kinds of situations, some of which are more or less related, but none identical (say) to the problem at hand, into precise quantitative specifications as needed for Jaynes' approach in an objective way, all before seeing the data. Setting aside the difficulties or working with informally specified prior information, even the more elementary key issue of specifying an objective prior distribution formalizing the absence of information is riddled with difficulties, and there are various principles for doing this which disagree in many cases (Kass and Wasserman, 1996). Objectivity seems to be an ambition rather than a description of what indeed can be achieved by setting up objectivist Bayesian priors. More modestly, therefore, Bernardo (1979) spoke of "reference priors," avoiding the term "objective," and emphasizing that it would be desirable to have a convention for such cases (O2b), but admitting that it may not be possible to prove any general approach for arriving at such a convention uniquely correct or optimal in any rational sense.

Apart from the issue of the objectivity of the specification of the prior, by and large the objectivist Bayesian approach has similar advantages and disadvantages regarding our list of virtues as the subjectivist Bayesian approach. Particularly it comes with the same difficulties regarding the issue of falsifiability from observations. Prior probabilities are connected to logical analysis of the situation rather than to betting rates for future observations as in de Finetti's subjectivist approach, which makes the connection of objectivist Bayesian prior probabilities to observations even weaker than in the subjectivist Bayesian approach (but probabilistic logic has applications other than statistical data analysis, for which this may not be a problem). The merit of objectivist Bayesianism is that the approach comes with a much stronger drive to justify prior distributions in a transparent way using principles that are as clear and general as possible.

This drive, together with some subjectivist honesty about the fact that despite trying hard in the vast majority of applications the resulting prior will not deserve the "objectivity" stamp and will still be subject to potential disagreement, can potentially combine the best of both of these traditional Bayesian worlds.

FalsificationistBayesianism

For both subjectivist and objectivist Bayesians, following de Finetti (1974) and Jaynes (2003), probability models including both parameter priors and sampling models do not model the data generating process, but rather represent plausibility or belief from a certain point of view. Plausibility and belief models can be modified by data in ways that are specified a priori, but they cannot be falsified by data.

In much applied Bayesian work, on the other hand, the sampling model is interpreted, explicitly or implicitly, as representing the data-generating process in a frequentist or similar way, and parameter priors and posteriors are interpreted as giving information about what is known about the "true" parameter values. It has been argued that such work does not directly run counter to the subjectivist or objectivist philosophy, because the "true parameter values" can often be interpreted as expected large sample functions given the prior model (Bernardo and Smith, 1994), but the way in which classical subjectivist or objectivist statistical data analysis is determined by the untestable prior assignments is seen as unsatisfactory by many statisticians.

The suggestion of testing aspects of the prior distribution by observations using error statistical techniques has been around for some time (Box, 1980). Gelman and Shalizi (2013) incorporate this in an outline of what we refer to here as "falsificationistBayesianism," a philosophy that openly deviates from both objectivist and subjectivist Bayesianism, integrating Bayesian methodology with an interpretation of probability that can be seen as frequentist in a wide sense and with an error statistical approach to testing assumptions in a bid to improve Bayesian statistics regarding virtue O4b. FalsificationistBayesianism follows the frequentist interpretation of the probabilities formalized by the sampling model given a true parameter, so that these models can be tested using error statistical techniques. Gelman and Shalizi argue, as some frequentists do, that such models are idealizations and should not be believed to be literally true, but that the scientific process proceeds from simplified models through test and potential falsification by improving the models where they are found to be deficient.

This reflects certain attitudes of Jaynes (2003), with the difference that Jaynes generally considered probability models as derivable from constraints of a physical system, whereas Gelman and Shalizi focus on examples in social or network science which are not governed by simple physical laws and thus where one cannot in general derive probability distributions from first principles, so that "priors" (in the sense that we are using the term in this paper, encompassing both the data model and the parameter model) are more clearly subjective (Gelman and Henning, 2015).

A central issue for falsificationistBayesianism is the meaning and use of the parameter prior, which can have various interpretations, which gives falsificationistBayesianism a lot of flexibility for taking into account multiple perspectives, contexts, and aims (S1, S2a) but may be seen as a problem regarding clarity and unification (O1a, O2c). Frequentists may wonder whether a parameter prior is needed at all. Here are some potential benefits of incorporating a parameter prior:

• The parameter prior may formalize relevant prior information.

• The parameter prior may be a useful device for regularization.

• The parameter prior may formalize deliberately extreme points of view to explore sensitivity of the inference.

• The parameter prior may make transparent a point of view involved in an analysis.

• The parameter prior may facilitate a certain kind of behavior of the results that is connected to the aims of analysis (such as penalizing complexity or models on which it is difficult to act by giving them low prior weight).

• The Bayesian procedure involving a certain parameter prior may have better error statistical properties (such as the mean squared error of point estimates derived from the posterior) than a straightforward frequentist method, if such a method even exists.

• Often finding a Bayesian parameter prior which emulates a frequentist/error statistical method helps understanding the implications of the method. Here are some ways to interpret the parameter prior:

• The parameter prior may be interpreted in a frequentist way, as formalizing a more or less idealized data generating process generating parameter values. The "generated" parameter values may not be directly observable, but in some applications the idea of having, at least indirectly, a sample of several parameter values from the parameter prior makes sense ("empirical Bayes"). In many other applications the idea is that only a single parameter from the parameter prior is actually realized, which then gives rise to all the observed data. Even in these applications one could in principle postulate a data generating process behind the parameter, of which only one realization is observable, and only indirectly. This is a rather bold idealization, but frequentists are no strangers to such idealizations either. A similarly bold idealization would be to view "all kinds of potential studies with the (statistically) same parameter" as the relevant population, even if the studies are about different topics with different variables, in which case more realizations exist, but it is hard to view a specific study of interest as a "random draw" from such a population. If parameter priors are interpreted in this sense, they can actually be tested and falsified using error statistical methods; see Gelman, Meng and Stern (1996). In situations with only one parameter realization, the power of such tests is low, though, and any kind of severe corroboration will be hard to achieve. Also, if there is only a single realization of an idealized parameter distribution, the information in the parameter posterior seems to rely strongly on idealization (Gelman and Henning, 2015).

• If the quality of the inference is to be assessed by error statistical measures, the parameter prior may be seen as a purely technical device. In this case, however, the posterior distribution does not have a proper interpretation, and only well defined statistics with known error statistical properties such as the mean or mode of the parameter posterior should be interpreted.

• Assuming that frequentist probabilities from sampling models should be equal to the subjectivist or objectivist epistemic probabilities if it is known that the sampling model is true (which Lewis, 1980, called "the principal principle"), the parameter prior can still be interpreted as giving epistemic probabilities such as subjectivist betting rates, conditionally on the sampling model to hold, even if the sampling model is interpreted in a frequentist way.

The possibility of rejecting the sampling model based on the data will invalidate both coherence and Cox's axioms, so that the foundation for the resulting epistemic probabilities becomes rather shaky. This does not necessarily have to stop an individual from interpreting and using them as betting rates, though. Given such a variety of uses and meanings, it is crucial for applications of falsificationistBayesianism that the choice of the parameter prior is clearly explained and motivated, so transparency is central here as well as for the other varieties of Bayesian statistics. Overall, falsificationistBayesianism combines the virtue of error statistical falsifiability with the virtues listed above as "subjective," doing so via a flexibility that may be seen by some as problematic regarding clarity and unification (Gelman and Henning, 2015).

Examples

In conventional statistics, assumptions are commonly minimized. Classical statistics and econometrics is often framed in terms of robustness, with the goal being methods that work with minimal assumptions. But the decisions about what information to include and how to frame the model— these are typically buried, not stated formally as assumptions but just baldly stated: "Here is the analysis we did . . . ," sometimes with the statement or implication that these have a theoretical basis but typically with little clear connection between subject-matter theory and details of measurements. From the other perspective, Bayesian analyses are often boldly assumption-based but with the implication that these assumptions, being subjective, need no justification and cannot be checked from data.

We would like statistical practice, Bayesian and otherwise, to move toward more transparency, with an intellectual "paper trail" linking theory and data to models, and recognition of multiple perspectives in the information that is included in this paper trail and this model. In this section we show how we are trying to move in this direction in two of our recent research projects. We present these examples not as any sort of ideals but rather to demonstrate how we are grappling with these ideas and, in particular, the ways in which active awareness of the concepts of transparency, consensus, impartiality, correspondence to observable reality, multiple perspectives and context dependence is changing our applied work.

Objectivity to Subjectivity

In multiple fields, as also in the career guidance field, we are witnessing a change in paradigms, namely from objective to subjective and quantitative to qualitative. This change is startling. Conventionally, the definition of scientific inquiry itself is the pursuit of truth in as objective a manner as possible. The new paradigm forces us to question whether truth can be singular, fixed or absolute. It questions whether truth can ever be a positional (i.e., without a position or a frame of reference). Moreover, it makes us re-examine objectivity as a characteristic of scientific inquiry; for that matter, for any inquiry.

Is the new paradigm heretical? Surely, the word objective means that which is unbiased, unprejudiced, and uncoloured by a particular frame of reference? Is it not desirable to see and act without a bias or prejudice? The immediate and conventional answer to these questions is yes. However if we pause and reflect, the answer, quite reasonably, is a no. Why? As on closer reflection, words such as bias are not negative in their literal sense although they have been in their connotative sense. A bias is a leaning; it defines a particular position. Following this train of thought, to be objective is to be without any leaning, to be minus a position or minus a point of view. Clearly, objectivity is controvertible. Subjectivity in the new paradigm has shed the garb of negatively biased, prejudiced or coloured.

Subjectivity represents the tenability of viewership or action as (always) constituted of some frame of reference, some position, or a particular point of view. Subjectivity moves from an inadmissible and abhorred condition to a valued and indispensable condition.

In continuing our re-examination of objectivity and subjectivity, note the subject and object in the following sentence: The woman looked at the tree. The tree is the object, the woman is the subject. A parallel is drawn for research: that which is studied is the object; the scientist/researcher is the subject. To clarify further, the adjective "objective" in the changing perspective is defined as that which comes from the object; that which belongs to the object. Complementarily, the adjective "subjective" is defined as that which comes from the subject; that which belongs to the subject. To reiterate, remember that the subject is the scientist or the researcher. Now ask yourself: In a scientific or any other inquiry, is knowledge produced by the object or the subject? Who is the direct producer of knowledge—the object or the subject?

If it is argued that it is the object that "reveals" its attributes, to be faithfully copied by the subject (see Piaget, 1983), why do we have multiple theories about the same phenomenon? It is because human knowledge is fundamentally creative and therefore plural. As Piaget (1983) has said, "knowledge results from interactions between the subject and the object, which are richer than what the objects provide by themselves;" it is "this inventive construction, which characterizes all living thought".

2.3 Identification of Research Problem

Research, in a very general term, is a systematic way for finding things you and other people did not know, which are called as research problems. In this sense, as what the discussion on the research process in previous modules has indicated, research is a process consisting of the identifying and defining research problem, formulating and testing the hypothesis through data collection, organization and analysis, making deductions and reaching of conclusion from the test results of the hypotheses, and reporting and evaluating the research.

Viewing its process, research is essentially a problem driven activity. Since research is problem driven, the first thing to deal in undertaking a study is to identify and determine the problem to study. Identifying a research problem is important because, as the issue or concern in a particular setting that motivates and guides the need for conducting a study, it lays the foundation for an entire project. If the foundation is shaky the entire project is doomed to failure. This is why novice researchers necessitate absolute caution in the initial stages of a research project. Professional researchers could easily identify a research problem because they have been quite familiar with the phenomena in which a problem generally presents itself.

Research Problem

A research problem, as mentioned previously, is the issue being addressed in a study. The issue can be a difficulty or conflict to be eliminated; a condition to be improved; a concern to handle; a troubling question, a theoretical or practical controversy (or a gap) that exists in scholarly literature. A research problem helps in narrowing the topic down to something that is reasonable for conducting a study. Creswell, 2012) defined research problem as "a general educational issue, concern, or controversy addressed in research that narrows the topic" p. 60).

Continuing with the topic identified by Pardede (2018), he looked at the perception of secondary schools students and teachers of the use of Indonesian in EFL classes. The problem is that very little attention has been given to the issue of first language use in English classes in Indonesia so that there is no appropriate empirical data to prepare the ground for a more reasoned use of Indonesian in the English classroom. By researching this problem, such necessitated data could be obtained. Based on the topic she identified, Angelianawati (2012) focused on the contribution of students' beliefs about language learning, learning styles, and language learning strategies on students' English achievement as the problem of her study. By knowing the role of those students' beliefs it would be easier to facilitate them to succeed their learning.

Locating the research problem in a research could be accomplished by asking ourselves the following questions.

(1) What was the issue, problem, or controversy that the researcher wanted to address?

(2) What controversy leads to a need for this study?

(3) What was the concern being addressed "behind" this study?

(4) Is there a sentence like "The problem addressed in this study is . . ."? (Creswell, 2012, p. 59).

The "problems" addressed in a study are usually stated at the end of the introduction or the literature review section. Some research article include them in a passage called the "statement of the problem" or in a paragraph put at the end of the introduction or the literature review section. Pardede (2018) stated his research problem at the final paragraph of the literature review. He wrote: "The problem addressed in this study is the perception students and English teachers towards the use of Indonesian in English classrooms at senior high schools around Jabodebek (Jakarta, Bogor, Depok, and Bekasi). Angelianawati, (2012) stated her research problem by writing "Therefore, this research was aimed at investigating how senior high schools students' beliefs about English language learning, their learning styles, and language learning strategies contributed to their English achievement" at the end of the final paragraph of the introduction section.

To better understand research problems, Creswell (2012, pp. 59-60) suggested to compare it to other parts of the research process, i.e. research topic, purpose, and research questions. After looking at their differences, you will see that they differ in terms of breadth from broad (topic) to narrow (specific research questions).According to him, as it has been previously mentioned, a research topic is the "broad subject matter addressed by the study." Pardede's (2018), research topic, for instance, is "the use mother tongue in foreign language learning". A research problem is "a general educational issue, concern, or controversy addressed in research that narrows the topic. The problem Pardede (2018) addressed is the perception students and English teachers towards the use of Indonesian in English classrooms at senior high schools around Jabodebek (Jakarta, Bogor, Depok, and Bekasi).

Steps to Defining the Research Problem

- Narrow the general topic down
- Literature review usually limited at this point
- Must make wise choices about what to investigate, study, explore
- Is the topic better suited to a qualitative or quantitative paradigm?
- Nature of the problem Previously studied, much literature – quantitative
- Exploratory study, lacking theory base – qualitative

Format of problem

- Statement – how you state the problem
- Question – implies relationship between two or more variables
- Statement – describes the scope of your work
- Hypothesis -- relationships
- Objective – achieve, measure

Clear & feasible problem statement

- Can it be understood by others?
- Can you describe it concisely, clearly?
- Do you demonstrate understanding of the area being investigated, studied?

Balance between general & specific in problem statement

- Avoid trivial problems that are meaningless
- Broad enough to be significant according to the criteria you establish
- Specific enough to be feasible for the research situation

2.4 Hypothesis Building

Often, one of the trickiest parts of designing and writing up any research paper is writing the hypothesis. A hypothesis is a tentative statement about the relationship between two or more variables. It is a specific, testable prediction about what you expect to happen in a study. It predicts the relationship between two or more variables, but it involves more than a guess. Most of the time, the hypothesis begins with a question which is then explored through background research. It is only at this point that researchers begin to develop a testable hypothesis.

Definitions of hypothesis

- "Hypotheses are single tentative guesses, good hunches – assumed for use in devising theory or planning experiments intended to be given a direct experimental test when possible". (Eric Rogers, 1966)

- "A hypothesis is a conjectural statement of the relation between two or more variables". (Kerlinger, 1956)

- "Hypothesis is a formal statement that presents the expected relationship between an independent and dependent variable."(Creswell, 1994)

- "A research question is essentially a hypothesis asked in the form of a question."

- "It is a tentative prediction about the nature of the relationship between two or more variables."

- "A hypothesis can be defined as a tentative explanation of the research problem, a possible outcome of the research, or an educated guess about the research outcome." (Sarantakos, 1993: 1991)

- "Hypotheses are always in declarative sentence form, as they relate, either generally or specifically, variables to variables."

- "An hypothesis is a statement or explanation that is suggested by knowledge or observation but has not, yet, been proved or disproved." (Macleod Clark J and Hockey L, 1981)

A hypothesis does not have to be correct. While the hypothesis predicts what the researchers expect to see, the goal of the research is to determine whether this guess is right or wrong. When conducting an experiment, researchers might explore a number of factors to determine which ones might contribute to the ultimate outcome.

In many cases, researchers may find that the results of an experiment do not support the original hypothesis. When writing up these results, the researchers might suggest other options that should be explored in future studies.

4.1. How Do Researchers Come up With a Hypothesis?

In many cases, researchers might draw a hypothesis from a specific theory or build on previous research. For example, prior research has shown that the people of mountainous area become hard-working. So a researcher might for a specific hypothesis that: Students of Shimla will acquire more marks in examination compared to the students of plain land.

4.2. Elements of a Good Hypothesis

When trying to come up with a good hypothesis for your own research or experiments, ask yourself the following questions:

- ❖ Is your hypothesis based on your research on a topic?
- ❖ Can your hypothesis be tested?
- ❖ Does your hypothesis include independent and dependent variables?

Before you come up with a specific hypothesis, spend some time doing background research on your topic. Once you have completed a literature review, start thinking about potential questions you still have. Pay attention to the discussion section in the journal articles you read. Many authors will suggest questions that still need to be explored.

An Example…

Imagine the following situation: You are a nutritionist working in a zoo, and one of your responsibilities is to develop a menu plan for the group of monkeys. In order to get all the vitamins they need, the monkeys have to be given fresh leaves as part of their diet. Choices you consider include leaves of the following species: (a) A (b) B (c) C (d) D and (e) E. You know that in the wild the monkeys eat mainly B leaves, but you suspect that this could be because they are safe whilst feeding in B trees, whereas eating any of the other species would make them vulnerable to predation. You design an experiment to find out which type of leaf the monkeys actually like best: You offer the monkeys all five types of leaves in equal quantities, and observe what they eat.

There are many different experimental hypotheses you could formulate for the monkey study. For example: When offered all five types of leaves, the monkeys will preferentially feed on B leaves. This statement satisfies both criteria for experimental hypotheses. It is a

•Prediction: It predicts the anticipated outcome of the experiment

•Testable: Once you have collected and evaluated your data (i.e. observations of what the monkeys eat when all five types of leaves are offered), you know whether or not they ate more B leaves than the other types.

Incorrect hypotheses would include:

When offered all five types of leaves, the monkeys will preferentially eat the type they like best. This statement certainly sounds predictive, but it does not satisfy the second criterion: there is no way you can test whether it is true once you have the results of your study. Your data will show you whether the monkeys preferred one type of leaf, but not why they preferred it (i.e., they like it best). I would, in fact, regard the above statement as an assumption that is inherent in the design of this experiment, rather than as a hypothesis.

When offered all five types of leaves, the monkeys will preferentially eat B leaves because they can eat these safely in their natural habitat.

This statement is problematic because its second part ('because they can eat these safely in their natural habitat') also fails to satisfy the criterion of testability. You can tell whether the monkeys preferentially eat baobab leaves, but the results of this experiment cannot tell you why. In their natural habitat, howler monkeys that feed in B trees are less vulnerable to predation than monkeys that feed on A, C, D, or E.

This is a perfectly good experimental hypothesis, but not for the experiment described in the question. You could use this hypothesis if you did a study in the wild looking at how many monkeys get killed by predators whilst feeding on the leaves of A, B etc. However, for the experimental feeding study in the zoo it is neither a prediction nor testable.

When offered all five types of leaves, which type will the monkeys eat preferentially? This is a question, and questions fail to satisfy criterion #1: They are not predictive statements. Hence, a question is not a hypothesis.

Types of Hypotheses

- NULL HYPOTHESES Designated by: H0 or HN Pronounced as "H oh" or "H-null"
- ALTERNATIVE HYPOTHESES Designated by: H1 or HA

The null hypothesis represents a theory that has been put forward, either because it is believed to be true or because it is to be used as a basis for argument, but has not been proved.

The alternative hypothesis is a statement of what a hypothesis test is set up to establish.

- Opposite of Null Hypothesis.
- Only reached if H0 is rejected.
- Frequently "alternative" is actual desired conclusion of the researcher!

Example

In a clinical trial of a new drug, the null hypothesis might be that the new drug is no better, on average, than the current drug. We would write H0: there is no difference between the two drugs on average.

The alternative hypothesis might be that: the new drug has a different effect, on average, compared to that of the current drug. We would write H1: the two drugs have different effects, on average. the new drug is better, on average, than the current drug.

We would write H1: the new drug is better than the current drug, on average.

We give special consideration to the null hypothesis… *f*

This is due to the fact that the null hypothesis relates to the statement being tested, whereas the alternative hypothesis relates to the statement to be accepted if / when the null is rejected.

The final conclusion, once the test has been carried out, is always given in terms of the null hypothesis. We either 'reject H0 in favor of H1' or 'do not reject H0'; we never conclude 'reject H1', or even 'accept H1'.

If we conclude 'do not reject H0', this does not necessarily mean that the null hypothesis is true, it only suggests that there is not sufficient evidence against H0 in favour of H1; rejecting the null hypothesis then, suggests that the alternative hypothesis may be true.

Formulating Hypothesis through Inductive and Deductive Approaches

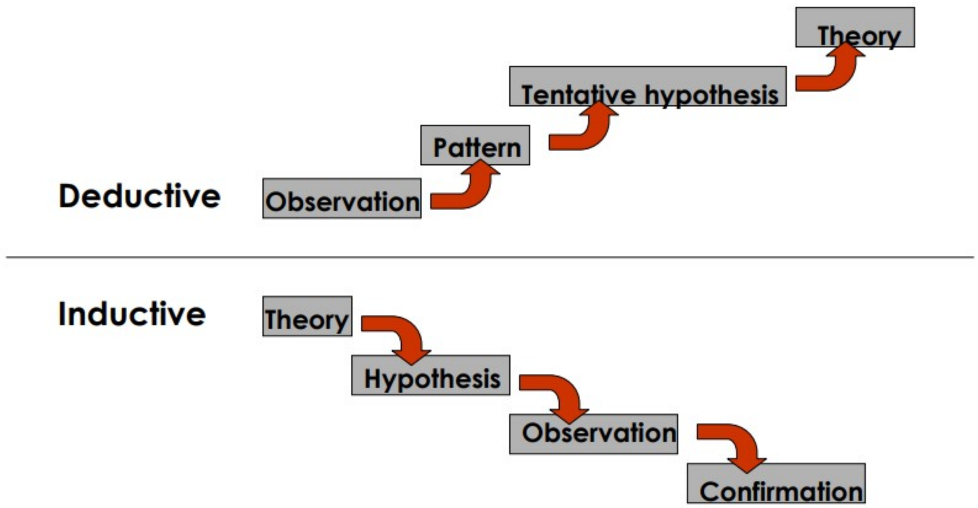


Figure: Hypothesis Building through Inductive and Deductive Approaches

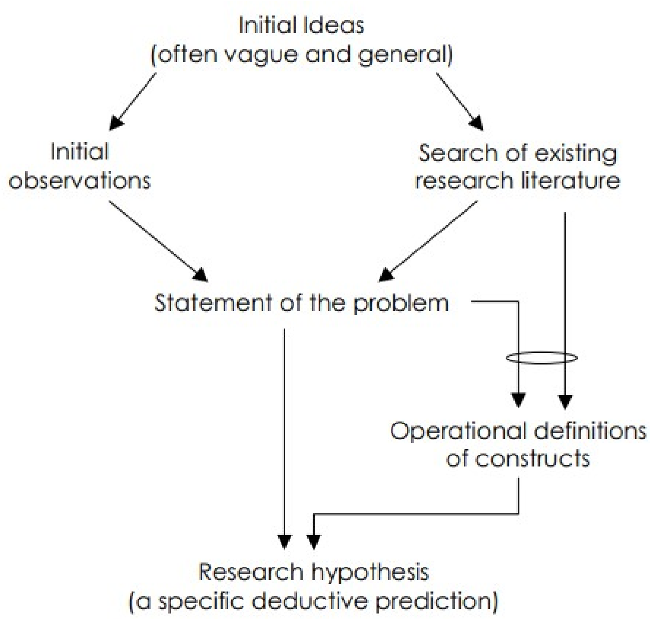


Figure: Generation of Research Hypothesis

Example:

Consider the example of a simple association between two variables, Y and X.

1. Y and X are associated (or, there is an association between Y and X).
2. Y is related to X (or, Y is dependent on X).
3. As X increases, Y decreases (or, increases in values of X appear to effect reduction in values of Y).

The first hypothesis provides a simple statement of association between Y and X. Nothing is indicated about the association that would allow the researcher to determine which variable, Y or X, would tend to cause the other variable to change in value. *f*

The second hypothesis is also a simple statement of association between Y and X, but this time it may be inferred that values of Y are in some way contingent upon the condition of the X variable.

The third hypothesis is the most specific of the three. Not only does it say that Y and X are related and that Y is dependent on X for its value, but it also reveals something more about the nature of the association between the two variables.

Testing & Challenging

The degree of challenge to the hypothesis will depend on the type of problem and its importance. It can range from just seeking "a good enough" solution to a much more rigorous challenge.

The term "challenging" may include *f*

- ❖ Verification
- ❖ Justification
- ❖ Refutability
- ❖ Validity
- ❖ Rectification
- ❖ Repeatability
- ❖ Falsification

Falsifiability

In the scientific method, falsifiability is an important part of any valid hypothesis. In order to test a claim scientifically, it must be possible that the claim could also be proven false. One of the hallmarks of a pseudoscience is that it makes claims that cannot be refuted or proven false. Students sometimes confuse the idea of falsifiability with the idea that it means that something is false, which is not the case. What falsifiability means is that if something was false, then it is possible to demonstrate that it is false.

There are two possibilities

1. Nothing Happened the Null Hypothesis - Ho
2. Something Happened the Alternative Hypothesis – H

Hypothesis testing is a four-step procedure:

1. Stating the hypothesis (Null or Alternative)

2. Setting the criteria for a decision
3. Collecting data
4. Evaluate the Null hypothesis

Two types of mistakes are possible while testing the hypotheses.

- Type I
- Type

Type I Error *ƒ*

A type I error occurs when the null hypothesis (H0) is wrongly rejected. For example, A type I error would occur if we concluded that the two drugs produced different effects when in fact there was no difference between them.

Type II Error *ƒ*

A type II error occurs when the null hypothesis H0, is not rejected when it is in fact false. For example: A type II error would occur if it were concluded that the two drugs produced the same effect, that is, there is no difference between the two drugs on average, when in fact they produced different ones.

Working Hypothesis

A working hypothesis is a hypothesis that is provisionally accepted as a basis for further research in the hope that a tenable theory will be produced, even if the hypothesis ultimately fails. Like all hypotheses, a working hypothesis is constructed as a statement of expectations, which can be linked to the exploratory research purpose in empirical investigation. Working hypotheses are often used as a conceptual framework in qualitative research.

The provisional nature of working hypotheses makes them useful as an organizing device in applied research. Here they act like a useful guide to address problems that are still in a formative phase.

In recent years, philosophers of science have tried to integrate the various approaches to evaluating hypotheses, and the scientific method in general, to form a more complete system that integrates the individual concerns of each approach. Notably, ImreLakatos and Paul Feyerabend, Karl Popper's colleague and student, respectively, have produced novel attempts at such a synthesis.

2.5 Methods of Sampling and Sample Design

When undertaking a research project, a researcher has a question or set of questions that hoe or she wants to answer. For example, if the researcher would like to know the information technology needs of University students in Thailand, she would have to identify the 'population'. In this case, the population would be all 'university students' at in Thailand. This illustrates a basic problem that all researcher' have to face in their research project. There are hundreds of thousands of 'university students' in Thailand?

Clearly, the researcher cannot study every university students in Thailand. This is where sampling comes into the research process. Sampling can be defined as "….the process of selecting units (e.g., people, organizations) from a population of interest so that by studying the sample we may fairly generalize our results back to the population from which they were chosen (Trachoma, 2006, n.p.)." It is possible therefore to use sampling techniques to select a smaller group - or sample - from the population that will statistically represent the whole population. It is often necessary to use sampling because researchers usually do not have the time, energy, money or resources to study the whole population.

The small group chosen to study and representing the total population is called a sample and the whole group which it is believed to represent is called universe. The number of individuals or observations in the sample is denoted as sample size.

Why sampling? to study the whole population?

A major reason studying samples rather than the whole group is that the whole group is so large that studying it is not feasible.

Example- college students in CA. If we can study the whole population, we do not need to go through the sampling procedures. Much research is based on samples of people.

Representativeness - how representative the selected cases are?

Then, can knowledge gained from selected cases be considered knowledge about a whole group of people? The answer depends on whether those selected cases are representative of larger group. News magazine articles about public opinion: How can we be sure that the results reflect the public's true opinion, in other words, how much they can represent views of all Americans? The ultimate purpose of sampling is to get accurate representativeness. The important consideration about samples is how representative they are of the population from which we draw them.

Casual vs. scientific sampling

In both daily life and practice, we are involved in sampling decisions - movies, car purchases, class selections, etc. to get feedbacks about service satisfaction from clients – what is said in community or agency meeting. How much of this information is representative? The information can be misleading or biased - The people who aLend or are the most vocal at a meeting may be the most satisfied (or most dissatisfied). If a sample size is too small, information can be biased as well. Scientific sampling is considerably more careful and systemic than casual, everyday sampling.

In research, scien5fic sampling procedures have been developed so that we can minimize the likelihood that samples we select will be biased or too small.

The adequacy of sampling frameis crucial in determining the quality of the sample. In other words, the degree to which the sampling frame includes all members of the popula5on is most important. Some of the adequate sampling frames consist of lists of members of organiza5ons. For example, membership rosters of NASW, APA.

The population consists of the sampling frame, and we can make legi5mate generaliza5ons only about the sampling frame. Many social workers, for example, do not belong to NASW. Thus, a sample taken from the NASW membership roster represents only NASW members and not all social workers. It is important to assess carefully who the list includes and who the list excludes.

5.1. Types of Sampling

A. Probability Sampling

- Simple random sampling
- Systematic sampling
- Stratified sampling
- Quota sampling

- Cluster sampling and area sampling.
- Multistage sampling
- Sequential sampling

B. Non-Probability Sampling

- Snowball sampling
- Convenience sample
- Judgment sample
- Quota sample

A. <u>Probability Sampling</u>

a) **<u>Simple random sampling</u>**: each member of the population has an equal chance of being selected in the sample.

Why Random Sampling?

- The word random refers to a process that generates a mathematically random result, one in which no humanly generated pattern exists.

- Social work researchers usually try to select their cases using a random procedure in order to assure that no human bias exists in the selection process. They hope that the inferences they draw from their study will be maximally generalizable, statistically accurate, and useful. Using random procedures allows the use of probability sampling methods.

Characteristics

- Applies random selection using random numbers
- Generally assumed in all probability sampling applications.
- Once a sampling frame has been established, the researcher assigns a single number to each member in the list without skipping any number in the process.
- A table of random numbers is sometimes used to select element for the sample
- Many computer programs can generate a series of random numbers.
- SRS is easy to accomplish and explain to others.
- SRS often is impractical, especially not most statistically efficient when dealing with large scale projects.
- Simple vs. not most efficient & not good representation of subgroups

b) <u>Stratified Random Sampling</u>:

Is applied when population from which sample to be drawn from the group does not have homogeneous group of stratified sampling technique, in generally it is used to obtain a representative of a good sample. Stratified type of sampling divide the universe into several sub group of population that are individually more homogeneous than the total population (the sub-populations differences are called strata) and select items will be selected from each stratum to generate a sample in this case each of the stratum will be more homogeneous with the population, more precise estimate will be generated from each for stratum (Etikan and Bala, 2017).

We get the estimate of the population from each stratum when there is better accuracy from each of the component; we get a better estimate of the whole. The stratified sampling gives more reliable and detailed information about the sample. The forming of strata is informed of purposive system from a well experience and special judgment of a researcher. The strata are defined by the population characteristics of the estimate. The fitted organized design for stratification is the pilot study, which assists in the determination of more appropriate and efficient planning for stratification and element within both of the stratum are homogeneous while element between each strata is heterogeneous.

Items selection from each separately stratum is done by using simple random sampling and systematic random sampling because they are reflected more proper in a convinced situations. Proportional allocation is used when the sample size from different stratum will be kept proportional to the strata size. To compare the difference for the strata, selecting equal sample from each of the stratum would be more efficient even though the strata will be different in sizes.

In cases the strata differs not only by size but also in variability and it is considered reasonable to take larger samples from the more variable of strata and smaller samples from the less variable strata and account for both differences of stratum size and differences of stratum variability. Disproportionate sampling design is required in the sample sizes of k strata which is called the optimum allocation.

- ✓ Strata**-** the population is divided into a number of sections.
- ✓ A sample is drawn independently from each stratum by the simple random sampling methods.
- ✓ A random sample is taken from each stratum the whole population is adequately represented**.**

c) <u>Systemic sampling</u>:

Every nth case is chosen for the study from a list of cases.

Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are to be selected by a fixed period, it is not like a random sample in real sense, systematic sampling has confident points of having improvement over the simple random sample, as ample the systematic sample is feast more equally completed to the complete population. The execution of the method is very easy, less in cost and conveniently to use in case of a larger population (Etikan and Bala, 2017).

Ex- Suppose we wish to take a sample of 20 cases from the 500 Hb. All the 500 values are arranged in some order and one observation is taken from each of 25 observations systematically. The first value to be selected among the first 25 observations to be determined from the table of random numbers. Suppose it is 10. Then our sample consists of the 10th value, 35th, 60th, 85th, etc. and the last one is 485th value.

d) <u>Cluster sampling</u>:

When the whole population consists of many natural groups

When the total area of the research is too large a better way for the researcher is to divide the area in to smaller part of the same or equal and then select randomly from the smaller units. it is expected that that the total population is to be divided in to relatively a smaller number which are still from the clusters of smaller units and then some of this cluster unit will be selected randomly so that it will be included in the general sampling. One of the advantages of using the cluster sampling is economical in reducing cost by concentrating on the selected clusters it gives less precision than the simple random sampling.

Ex- In an immunization status survey of the first standard children in a town a child may be regarded as a sampling unit and a random sample of the required number of children can be chosen using the normal rolls kept in each school. Alternatively the school itself can be considered as the sampling unit and random sample of the schools selected. Thus all the first standard children in these selected schools are included in the sample.

Non-Probability Sampling

Non-probability sampling is a sampling procedure that will not bid a basis for any opinion of probability that elements in the universe will have a chance to be included in the study sample. We are going to see from diverse method of five different sampling considering the non-random designs. which are; Quota sampling, Accidental sampling, Judgemental sampling or Purposive sampling, Expert sampling, Snowball sampling, Modal instant sampling .From the listed the researcher has to deliberately select items to be sample. This type of sampling is costly in application.

a) **Snowball sampling**: Participants area reached through contacts from one to many

It isa design process of selection usually done by using, networks. It is useful when the researcher know little about a group or organisation to study; contact with few individuals will direct him to other group. The selection of the study sample will be useful for communication aspect, in making decision or in diffusion of knowledge to people. The disadvantage is that the choice of the whole sample balances on the choice of individuals from the beginning of the stage, belonging to a particular clique or have ample biases. It will difficult to use when the sample becomes larger and larger.

In some communities (especially those in developing countries), the only feasible way to find its members is by asking other members. The first step in this procedure is to find a few members of the population using any method. This step is denoted as the first round. Then you ask each of these first-round members if they know of any others. The names given will form the second round. Then you go to each of those second-round people, and ask them for more names. This process is repeated for several more rounds. The process is stopped when you start hearing about the same people over and over again. The methodology used to stop the procedure can be described.

For each round, count the number of names you get and the number of new names obtained in this round. Then calculate the percentage of new names to the total number of names. For example, if the second round gives you 100 names, but 30 of them are for people who were mentioned in the first round, then the percentage of new names for that round is 70 per cent. This percentage of new names is high at first, but then drops sharply.

When the percentage of new names drops to around 10 per cent, then stop. This often happens at the fourth or fifth round. After performing this, something close to a list of the whole population is available, and many of the population members will know that you are planning some research. Using that list, you can draw a random sample. Snowball sampling works well when members of a population know each other. The problem with snowball sampling is that isolated people will not be included in the study.

b) **Convenience sample**: The researcher selects the easiest population members from which to obtain information.

c) **Judgment sample**:

The researcher uses his/her judgement to select population members who are good prospects for accurate information. The sampling design is based on the judgement of the researcher as to who will provide the best information to succeed for the objectives study. The person conducting the research need to focus on those people with the same opinion to have the required information and be willing of sharing it.

d) **Quota sample**:

The researcher finds and interviews a prescribed number of people in each of several categories. The researcher here is ease of access to his sample population by using quota sample, his tallying will be at his convenience guide by some evident of characteristic, such as sex, race, based on population of interest. The sample selection is by the convenient door of the researcher, any person or individual mistakenly seen with the same characteristics will be asked pertaining the subject of the research for inclusion. It will flow in the same manner until the desired number is achieved.

Quota sampling is of two types; first proportionate quota sampling represent the characteristics of major population by sampling a proportional total. Example if we are interested in studying population of 40 per cent of females and 60 per cent of males. We need a 100 size for the sample; the selection will not stop unless the target is hit before stopping. Meanwhile when the exact number of either male or female is gotten, say 40 female, the selection for the male has to continue in the same process, eventually when a legitimate female comes across, it will not be selected because there number is already completed. The major setback of purposive sampling is that you necessity to agree on the specific features of the quota to base on. This will be either to base on religion, age, education gender; etc.

The non-proportional quota sampling is a technique with small restriction of minimum of sample number of unit from each category. It's not interested in having a number that will match the proportions of the population. Rather need to have sufficient to guarantee that you will be capable to talk about even a small cluster in the population. The method is a non-probabilistic sampling that typically used in assuring that small groups of samples are adequately represented.

Sampling errors

The estimates from a sample survey are affected by two types of errors:

1) non-sampling errors, and

2) sampling errors.

Non-sampling errors are the results of mistakes made in implementing data collection and data processing, such as failure to locate and interview the correct household, misunderstanding of the questions on the part of either the interviewer or the respondent, and data entry errors. To minimize this type of error, numerous efforts have to be made during the implementation of the survey. These efforts include high quality and intensive training, good field supervision, double entry of the data to minimize data entry errors. However, non-sampling errors are impossible to avoid and difficult to evaluate statistically (Elder, 2009).

Sampling errors, on the other hand, can be evaluated statistically. The sample of respondents selected in the SWTS survey is one of many samples that could have been selected from the same population, using the same design and expected size. Each of these samples would yield results that differ somewhat from the results of the actual sample selected. Sampling errors are a measure of the variability between all possible samples. Although the degree of variability is not known exactly, it can be estimated from the survey results.

A sampling error is usually measured in terms of the standard error for a particular statistic (mean, percentage, etc.), which is the square root of the variance. The standard error can be used to calculate confidence intervals within which the true value for the population can reasonably be assumed to fall. For example, for any given statistic calculated from a sample survey, the value of that statistic will fall within a range of plus or minus two times the standard error of that statistic in 95 per cent of all possible samples of identical size and design.

If the sample of youth had been selected as a simple random sample, it would have been possible to use straightforward equations for calculating sampling errors. However, a multi-stage stratified design needs to use more complex equations. To calculate sampling errors for the SWTS, if it is a multi-stage sample, one has to use the Taylor linearization method of variance estimation for survey estimates that are means or proportions.

2.6 Methodological Orientation: Quantitative and Qualitative

Qualitative research is characterised by its aims, which relate to understanding some aspect of social life, and its methods which (in general) generate words, rather than numbers, as data for analysis.

➢ Quantitative Methodologies

Quantitative methods emphasize on objective measurements and numerical analysis of data collected through polls, questionnaires or surveys.It requires larger sample size for drawing conclusions and for generalisation of findings on the target population across different groups and communities. Hence, it is essential to scientifically explain the sample plan, design of tools, the validity of data and statistical applications.

Quantitative research is an approach for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures. The final written report has a set structure consisting of introduction, literature and theory, methods, results, and discussion. Like qualitative researchers, those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings.

Advantages of Quantitative Methods

- A big advantage of this approach is that the results are valid, reliable and generalizable to a larger population.
- Quantitative research is advantageous for studies that involve numbers, such as measuring achievement gaps between different groups of students or assessing the effectiveness of a new blood pressure medication.

Limitations of Quantitative Methods

- The in-depth analysis is hardly possible due to its predominant empirical nature,
- Due to ‘structural bias’, many times the interpretation and discussion actually reflect the views of the researcher instead of the participating subject
- It is not effective for any micro-level analysis of social and behavioural problems

➢ Qualitative Methodologies

Qualitative research is often used to conduct social and behavioural studies because human interactions are more complex than molecular reactions in a beaker. Subjectivity, non-random sampling and small sample size distinguishes qualitative research from quantitative research.

Qualitative research is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem. The process of research involves emerging questions and procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data. The final written report has a flexible structure. Those who engage in this form of inquiry support a way of looking at research that honours an inductive style, a focus on individual meaning, and the importance of rendering the complexity of a situation.

Advantages of Qualitative Research Methods

- ✓ Understand the perspectives of participants;
- ✓ Explore the meaning they give to phenomena;
- ✓ Observe a process in depth...
- ✓ A big advantage of qualitative research is the ability to deeply probe and obtain rich descriptive data about social phenomena

Limitations of Qualitative Research Methods

 Researchquality is heavily dependent on the researcher's skills and hence may not be free from researcher's personal biases;
 Issues of anonymity and confidentiality can present problems when presenting findings; and
 The process is often time consuming and findings cannot be generalised to a larger population.

Table: Comparison between Qualitative and Quantitative Research Methods

Bases of Comparison	Qualitative Research	Quantitative Research
Meaning	Qualitative research is a method of inquiry that develops understanding on human and social sciences, to find the way people think and feel.	Quantitative research is a research method that is used to generate numerical data and hard facts, by employing statistical, logical and mathematical technique.
Nature	Holistic	Particularistic
Approach	Subjective	Objective
Research type	Exploratory	Conclusive
Reasoning	Inductive	Deductive
Sampling	Purposive	Random
Data	Verbal	Measurable
Inquiry	Process-oriented	Result-oriented
Hypothesis	Generated	Tested
Elements of analysis	Words, pictures and objects	Numerical data
Objective	To explore and discover ideas used in the on-going processes.	To examine cause and effect relationship between variables.
Methods	Non-structured techniques like In-depth interviews, group discussions etc.	Structured techniques such as surveys, questionnaires and observations.
Result	Develops initial understanding	Recommends final course of action

2.7 Abstract, and Summary and Synopsis: Their Differences

An **abstract** is a brief summary of a research article, thesis, review, conference proceeding or any in-depth analysis of a particular subject or discipline, and is often used to help the reader quickly ascertain the paper's purpose. When used, an abstract always appears at the beginning of a manuscript, acting as the point-of-entry for any given scientific paper or patent application. Abstraction and indexing services are available for a number of academic disciplines, aimed at compiling a body of literature for that particular subject. An abstract is a self-contained, short, and powerful statement that describes a larger work. Components vary according to discipline; an abstract of a social science or scientific work may contain the scope, purpose, results, and contents of the work. An abstract of a humanities work may contain the thesis, background, and conclusion of the larger work. An abstract is not a review, nor does it evaluate the work being abstracted. While it contains key words found in the larger work, the abstract is an original document rather than an excerpted passage.

Why do we write abstracts?

Abstracts are important parts of reports and research papers and sometimes academic assignments. The abstract is often the last item that you write, but the first thing people read when they want to have a quick overview of the whole paper. We suggest you leave writing the abstract to the end, because you will have a clearer picture of all your findings and conclusions.

How do I write an abstract?

♣First re-read your paper/report for an overview. Then read each section and condense the information in each down to 1-2 sentences.

♣ Next read these sentences again to ensure that they cover the major points in your paper.

♣ Ensure you have written something for each of the key points outlined above for either the descriptive or informative abstract.

♣ Check the word length and further reduce your words if necessary by cutting out unnecessary words or rewriting some of the sentences into a single, more succinct sentence.

♣ Edit for flow and expression. What makes a good abstract? A good abstract:

♣uses one well-developed paragraph that is coherent and concise, and is able to stand alone as a unit of information

♣covers all the essential academic elements of the full-length paper, namely the background, purpose, focus, methods, results and conclusions

♣contains no information not included in the paper

What makes a good abstract?

A good abstract:

♣uses one well-developed paragraph that is coherent and concise, and is able to stand alone as a unit of information

♣covers all the essential academic elements of the full-length paper, namely the background, purpose, focus, methods, results and conclusions

♣contains no information not included in the paper

♣is written in plain English and is understandable to a wider audience, as well as to your discipline-specific audience

♣often uses passive structures in order to report on findings, focusing on the issues rather than people

♣uses the language of the original paper, often in a more simplified form for the more general reader

♣usually does not include any referencing

♣in publications such as journals, it is found at the beginning of the text, while in academic assignments, it is placed on a separate preliminary page.

(1) Descriptive abstracts

Descriptive abstracts are generally used for humanities and social science papers or psychology essays. This type of abstract is usually very short (50-100 words). Most descriptive abstracts have certain key parts in common. They are:

♣background
♣purpose
♣particular interest/focus of paper
♣overview of contents (not always included)

(2) Informative abstracts

Informative abstracts are generally used for science, engineering or psychology reports. You must get the essence of what your report is about, usually in about 200 words. Most informative abstracts also have key parts in common. Each of these parts might consist of 1-2 sentences. The parts include:

♣background
♣aim or purpose of research
♣method used
♣findings/results
♣conclusion

The table below summarises the main features of, as well as the differences between, the two types of abstracts discussed above. In both types of abstract, your lecturer/tutor may require other specific information to be included. Always follow your lecturer/tutor's instructions.

Descriptive abstract	*Informative abstract*
Describes the major points of the project to the reader. Includes the background, purpose and focus of the paper or article, but never the methods, results and conclusions, if it is a research paper. Is most likely used for humanities and social science papers or psychology essays.	Informs the audience of all essential points of the paper. Briefly summarises the background, purpose, focus, methods, results, findings and conclusions of the full-length paper. Is concise, usually 10% of the original paper length, often just one paragraph. Is most likely used for sciences, engineering or psychology reports.

An **executive summary** is a report, proposal, or portfolio, etc. in miniature. That is, the executive summary contains enough information for the readers to become acquainted with the full document without reading it. Usually, it contains a statement of the problem, some background information, a description of any alternatives, and the major conclusions. Someone reading an executive summary should get a good idea of main points of the document without becoming bogged down with details.

Preparing to Write

To write a good summary it is important to thoroughly understand the material you are working with. Here are some preliminary steps in writing a summary.

1. Skim the text, noting in your mind the subheadings. If there are no subheadings, try to divide the text into sections. Consider why you have been assigned the text. Try to determine what type of text you are dealing with. This can help you identify important information.

2. Read the text, highlighting important information and taking notes.

3. In your own words, write down the main points of each section.

4. Write down the key support points for the main topic, but do not include minor detail.

5. Go through the process again, making changes as appropriate.

When writing the summary there are three main requirements:

1. The summary should cover the original as a whole.

2. The material should be presented in a neutral fashion.

3. The summary should be a condensed version of the material, presented in your own words.

What a Summary Should Not Include

- Your opinion –
- What you think the author should have said
- Copied material or a string of quotes from the selection

For some of your Fast-Write homework assignments you'll be asked to summarize an article. Guidelines below will help guide your reading and writing:

1. Complete

A summary should include all the ideas that are essential to the author's thesis.

2. Concise

A summary should be considerably shorter than the passage. Do not include unessential information (length depends upon the purpose and your use of the summary. It could be one-half, one-third or one-eighth the length of the original.)

3. Accurate

A summary should represent the author's ideas. Do not distort the author's views.

4. Objective

A summary should recapitulate the author's points. Do not include your objections or criticisms in the summary.

5. Coherent

A summary should make sense to someone who has not read the original. It should not sound like a list of loosely-related sentences that have been strung together in paragraph format.

6. Independent

A summary should be written in your own words. Do not take strings of words from the source; do not paraphrase.

Difference between paraphrase and summary

***Paraphrase**: contains all the information in the source. No part of the original is left out. Writer rephrases original in his words.

****Summary**: contains only the most important information of the original. It does not have to follow the organization or order of the original. You do not change the meaning of the original, and you must give clear references as to the origin of the ideas.

How to Write a Summary

Being able to write a summary can prove that you were able to successfully understand a certain text to the point that you could easily write down a summary that contains nothing but your own words. Writing summaries can be tricky, this is why we have here steps on how to read a summary that will enable you to summarize texts effectively:

a) Divide your assigned reading into sections. Before getting straight to reading your assigned text, you have to make sure that you have divided your reading into sections so that you will be able to take in things slowly. Start off by focusing to the headings and subheadings first and that you would look at bold-faced terms as well. Understand these first so that you will be able to accomplish the following steps.

b) Read your assigned text. Now that you are able to prepare and divide your assigned text, you can now easily read your assigned text. Just read straight through and you do not have to stop in case something would trouble you—you just have to make sure that you will be able to get a feel for the tone, style, and main idea that the author tried to convey.

c) Reread your assigned text. On the second time that you will be reading your assigned text, you have to make sure that you will be able to read actively. This means that you should take note and underline the topic sentences and the key facts of the entire text. You may also take note and label the areas you want to discuss later on your summary writing. You may also take note and label the parts wherein you think your summary shouldn't contain or the ideas, even though interesting, that are too specific that it almost gives away everything the original text contains. It is also in rereading your text that you will be able to identify the parts of the original text that you were not able to comprehend when you were just skimming and reading the assigned text straight through.

d) Start writing by writing one sentence at a time. You have now thoroughly read your assigned text, you should now definitely have a firm grasp on how you are going to summarize your assigned text. From the first to the third steps, you have divided your assigned text into manageable texts, you can now easily write down the main idea of each of the sections from your assigned text. Knowing the main ideas or key points of your assigned text will enable you to write well-developed sentences for your summary.

e) Write a thesis statement. If you have always wondered what the best key to a well-written summary is, then now it's time for you to know that it's the thesis statement that makes up an excellently written summary. From the sentences you have written on the fourth step, it is possible for you to create a thesis statement that will be able to clearly communicate what the original text tried to convey and achieve. In the event that this step is difficult, it is best for you to go back to the previous steps since to get a thesis statement; you have to be able to understand the main points of the original text first.

f) Ready to write. At this point, you might wonder that you have written enough on the previous steps, but the actual writing had just begun. Your thesis writing can be used as the introductory sentence of your summary and what you will be doing in this step is to write the other sentences that will make up the rest of the body of your summary in which that you will be able to write in order. It will be effective if you will add transition words such as then, however, also, and moreover that can help you in creating an effective overall flow and structure of the summary that you are going to write. Keep in mind these tips as you do so:

- Always write in active voice and in the present tense.
- Always include the full name of the author and the full title of the work.
- Always keep things concise because the length of your summary should not equate to the original text.
- Cite the exact words of the author if you must use it.
- Always avoid adding your personal opinions, ideas, or interpretations into the summary that you are writing because the purpose of summary writing is to be able to accurately convey what the author's message is and not to provide criticism to his or her work (this is why you have to do a lot of reading).
- Revise what you have written. Even if you are very much certain about the accuracy of your summary, you would still have to revise what you have written. Check and revise it for style, grammar, and punctuation because even though these are the basic ones, there is still a possibility that you will be able to miss things out.

Major Attributes of a Summary

Now that you know the steps on how to create an idea, you should now learn what makes the major attributes that make up an effective summary:

- ❖ A summary should be comprehensive. Make your summary always comprehensive that the reader can definitely see that you have highlighted the major points from the assigned text and that this had been arranged well in a list. If you fail to do this, it would result in your summary looking like a complete chaos. Make sure that you will be able to meticulously choose only the ideas that could effortlessly explain what the author's thesis is.
- ❖ A summary should be concise. Keep in mind that you are to write a summary and not a reflection paper of what you have just so it is just right that you have to write a summary in a concise manner. Sometimes, it can be too hard to summarize a particular original text especially if there are so many points that you want to include. If this is your case, then you must make sure that you are only to take note of the main points, to avoid following the author's flow or pattern in the original text, and most especially, to avoid restating any similar ideas.

- A summary should be coherent. Sure, you are to restate the main points of the original text; however, you would also have to make sure that you will be able to connect each of these main points effectively in a way that it would be exactly the way the original text had conveyed. And sure you are to restate the main points of the original text with respect to what the original text conveyed, you should make sure that it could stand as a separate write-up or paper.

- A summary should be written independently. When writing a summary, it is not necessary for you to imitate how the author had summarized his or her own text. In fact, you should be able to write the summary using your own understanding, style, and manner. Restate it with the use of your own understanding. Do not even consider quoting the writer. However, even if it is highly suggested for you to write in your own words, you have to also make sure that you are not distorting the ideas and the intention of the writer of the original text that you are writing a summary out of.

A **synopsis** is a brief overview of a report's most important points, designed to give readers a quick preview of the contents. It's often included in long informational reports dealing with technical, professional or academic subjects and can also be called an abstract. Because it's a concise representation of the whole report, it may be distributed separately to a wide audience; interested readers can then order a copy of the entire report.

A synopsis is a summary. The purpose of a summary is to provide a shortened version of a givenpiece of writing. Most frequently a synopsis is a multi-paragraph summary of a chapter, book, article, or drama. It makes no effort to address every idea. The reader understands that a synopsis is necessarily subjective because the writer must select the items to be included. Thus, the synopsis reflects editorial bias.

Characteristics

A synopsis generally

- Selects main ideas which, in the writer's opinion, best represent the original piece
- Reflects the style of the current writer, as opposed to the author of the original
- Gives sufficient details to clarify the main ideas Is sufficiently informative to aid further investigation
- Has a reporter's objectivity

Process

Use the following process to develop a synopsis:

Step 1
Reading the Material

Prewriting

Reading the Material Quickly read the original passage without taking notes. Look for repeated themes and the general treatment of ideas. If unfamiliar names are difficult to you, you may want to list them for later reference, noting the correct spelling. Once you've finished reading, jot down the three or four most important ideas. They will serve as the skeletal outline of your synopsis. Remember, you will not cover all the ideas in a synopsis.

Step 2
Writing

Preparing the Synopsis Begin the synopsis by identifying in a single sentence the main idea of the original. Weave the title and author into an early (though not necessarily the first) sentence. Follow with a brief description of the main ideas. Follow the organization of the original. Keep in mind that a synopsis will be short, maybe only a paragraph in length. Deal only with essentials.

Step 3
Revising

Improving the Content Your synopsis will almost invariably be too long. Begin cutting. Omit needless details. Shorten sentences. Use strong nouns and verbs in order to omit modifiers. Keep cutting. Try this challenge: If you had to summarize the original in a single sentence, what would you say? If you could add on more sentences, what would you say? Follow this progression until you have stated all the important ideas.

Step 4
Proofreading

Checking the Details Once you have completed the revision, check for spelling, punctuation, grammar, mechanics, and usage. Refer also to the dictionary, a thesaurus, and/or the spell check/grammar check for spelling and vocabulary problems. The steps above should enable you to develop a synopsis. Compare your own with the example below, and study the analysis which follows.

Table: Differential Characteristics of Abstract, Executive Summary and Synopsis

Bases of Differences	Abstract	Executive Summary	Synopsis
Nature	Abbreviated summary, generally within 200-500 word limit	Unique selling point (USP).	A brief overview of a report's most important points, designed to give readers a quick preview of the contents
Audience	Specialized (researchers) or mere readers	Decision makers, e.g. corporate managers	technical, professional or academic
Scope	Informational, academic, administrative, and other general documents (thesis, articles, and patents)	Solicited or unsolicited sales proposals and bids (P&B)	
Purpose	**Give information** Ascertain the purpose of the whole document; give an overview or preview of its content.	**Call for action** Persuade readers to buy on their commended solution addressing the problem, namely, make your unique selling point (USP)	
Content	**Mainly technical**: 1.Present the problem and scope; 2. Expose the used methodology; 3. Report observations and results;	**Mainly managerial** 1.State outcomes and benefits; 2.Substantiate benefits with proofs of concept; 3.Apply benefits to the reader's particular; context (win	a concise representation of the whole report

	4. Draw conclusions and recommendations.	themes); 4.Recommend a solution to address the problem.	
Length	Short, Shorter than the executive summary.	Short, longer than the abstract	Synopsis is a prose table of contents that outlines the main points of the report.
Style	Technical, static, and more academic	Managerial, dynamic, and more enthusiastic	Less comprehensive than summary

2.8 Reference and Bibliography

A **bibliography** usually contains all the works cited in a paper, but it may also include other works that the author consulted, even if they are not mentioned in the text.

In **referencing**, each reference cited in text must appear in the reference list, and each entry in the reference list must be cited in text.

8.1. Referencing Style

There are different styles of referencing. Generally, American Psychological Association (APA) style is mostly followed in social sciences.

8.1.1. For Books

Authors surname (alphabetically), followed by their initials, Date of publication, Title of book in italics, Place of publication, Publisher. e.g.,

Doxiadis, C. A. (1968). *Ekistics: An Introduction to the Science of Human Settlements.* New York: Oxford University Press.

8.1.2. For Journal Article

The title of the article appears in inverted commas and name of the journal comes in italics, followed by volume number and pages of the article. e.g.

Bhagat, R. B. (2011). Emerging Pattern of Urbanization in India. *Economic & Political Weekly, XLVI (34)*, 10-12.

The terms 'References' and 'Bibliography' are often defined as having the same or similar meaning, consequently causing confusion. However, there is a difference in meaning between them.

REFERENCES are the items you have read and specifically referred to (or cited) in your assignment.

A **BIBLIOGRAPHY** is a list of everything you read in preparation for writing an assignment, whether or not you referred specifically to it in the assignment. A bibliography will, therefore, normally contain sources that you have cited and also those you found to be influential, but decided not to cite. A bibliography can give a tutor an overview of which authors have influenced your ideas and arguments even if you do not specifically refer to them.

What to Reference

You can cite references taken from a range of sources, e.g.

¬ Internet
¬ Notes supplied, and verbal comments made, by a learning tutor
¬ CD databases
¬ Books written by a single author
¬ Multiple edited books with contributions from a range of different authors
¬ Reference books of all types
¬ Legal documents
¬ Articles from journals
¬Reports of various kinds, e.g. official reports from government departments, university working papers, etc.
¬ Newspaper articles
¬ Papers presented at conferences
¬ Radio/TV/DVDs/audio cassette/CD Roms
¬ Interview transcripts
¬ Email correspondence
¬ Cinema films and theatre plays

In short, most information that has been written, recorded, filmed or presented in some way to others can potentially be used.

The important thing is to choose **reliable** sources that give credence, authority and support to the ideas and arguments that you present. Your tutor will suggest a range of reliable sources and this will be your starting point, but you will also be expected to look beyond the recommended reading and to search out relevant information for yourself.

You will find, however, that recommended books and other sources will prove – because of the accurate referencing that has gone into them – to be rich veins of additional information. If you read a particular chapter as a starting point for research into an assignment topic, often the references or bibliography will point you in the right direction of other relevant reading.

When to Reference

You should cite your sources of evidence in assignments in the following situations:

1. To give the reader the source of tables, photos, statistics and diagrams included in your assignment. These may be items directly copied or which have been a source of collation for you.

Example: In Britain, the proportion of employees on temporary contracts rose only marginally between 1992 and 1998, from 5.9 per cent to 7.4 per cent, and has since fallen to 7.1 per cent in 2000 (Office for National Statistics 2000).

2. When describing or discussing a theory, model or practice associated with a particular writer.

Example: The term 'instrumental or operant conditioning' is associated with the American Psychologist, B.F. Skinner (1956), and describes a process of shaping behaviour by a variety of means that encourage and reinforce desired behaviour or discourage unwanted behaviour.

3. To give weight or credibility to an argument presented by you, or supported by you, in your assignment.

Example: However, it can be argued that the corrosive social effects of workers having to manage increasing workloads outweigh these extrinsic advantages. Handy (1994, p.9), for example, suggests that businesses prefer to recruit "half as many people, paid twice as well and producing three times as much", with a destructive effect on the social lives of these core workers.

4. When giving emphasis to a particular idea that has found a measure of agreement and support amongst commentators.

Example: As the behavioural response of communication apprehension (CA) is to avoid or discourage interaction with others it is not surprising that CA has been linked to feelings of loneliness, isolation, low self-esteem and the inability to discuss personal problems with managers or others (Daly and Stafford, 1984; McCroskey, Daly, Richmond and Falcione, 1977; McCroskey and Richmond, 1987; Richmond, 1984; Scott and Rockwell, 1997).

(Note: The student cites five sources - all much saying the same thing - to emphasise and give credibility to an important point summarised in the assignment. The use of multiple authors can add weight to summary, particularly if the idea is a controversial one. However, citing six authors would be the maximum for this purpose, and citing two or three is a more usual practice).

5. To inform the reader of sources of direct quotations or definitions in your assignment.

Example

Pearson (1995) however, argues that a "search for a solution to ethical dilemmas using the methods of moral philosophy has failed" (p.3). He asserts that any approach to business ethics must take full account of the business perspective and an appreciation of business boundaries, albeit with account to the changing nature of these.

(Note: If the quote is taken from a printed book or journal, you always need to include the page number so the reader can go straight to that page to find it. Lengthy quotations (over two lines) should be indented in your assignment. This means you compress the quotation, italicise it and create a margin that distinguishes it from your own text, as per the example below. You don't need to use quotation marks in an indented quotation):

Example

of indented quotation: Robert Reich (2001) has argued that pay is proportionate to the skill you offer in the labour market: If you have been in a job that's rote or routine…or your job can be done by computerized machines or by software over the Internet – you're likely to be paid less than you used to be paid for doing it…(p.32)

6. To avoid plagiarism.

7. When paraphrasing another person's idea that you feel is particularly significant or likely to be a subject of debate; this can include definitions.

Example

We all perceive the world around us in ways that are often unique to us through a series of personal filters and we 'construct' our own versions of reality (Kelly 1955). (Note: In this example, the student paraphrases an idea that Kelly originally outlined in 1955. The inverted commas around 'construct' suggest this is a significant word used by Kelly to describe a key concept). By citing the source the student is, in effect, saying 'this is Kelly's idea; I am just paraphrasing it', and thus avoids accusations of plagiarism.

Plagiarism

One general definition of plagiarism is to knowingly take and use another person's work and, directly or indirectly, **claim it as your own**.

There are no internationally agreed academic norms or conventions on what constitutes plagiarism and this can cause difficulties for some international students who may have encountered different practices in their home countries.

In Britain, there is a particularly strong emphasis given to respecting the authorship of ideas and honouring the hard work that goes into researching, preparing and writing academic texts.

An academic monograph or textbook, for example, can take an author several years to research and write. Consequently it is widely felt in Britain that to copy from a book without acknowledging the source is a violation against the author's ownership of ideas and therefore morally wrong. For this reason plagiarism is treated seriously and blatantly plagiarised work is usually disqualified.

Each university develops its own interpretation of plagiarism, but in general there are four main forms of plagiarism:

- copying another person's work, including the work of another student (with or without their consent), and claiming or pretending it to be your own;
- presenting arguments that use a blend of your own and the copied words of the original author without acknowledging the source;
- paraphrasing another person's work, but not giving due acknowledgement to the original writer or organisation publishing the writing, including Internet sites;
- colluding with other students and submitting identical or near identical work.

However, it is also important that students are aware of their university's own interpretation of plagiarism, as each institution may place emphasis on a particular feature of plagiarism.

Of course, there will inevitably be some overlap between the writer's words and your own – particularly when describing places, dates, specific features and the names of organisations.

However, you should make a determined effort to use your own words to sum up what you have read. The act of doing this encourages a deeper level of understanding as, in the process, you are forced to think hard about what is actually said and meant by the authors.

Lecturers marking course work can recognise plagiarism easily. This applies particularly when passages are copied straight from books or cut and pasted from the Internet with no acknowledgement of their source. Lecturers will usually recognise the work of established writers in the subject area concerned and there will be stylistic differences in writing between the original author and a student's work that an experienced lecturer will detect. Any assignment maybe assessed through scanning software used by the Institute in order to gauge any level of plagiarism, the outcome of which will be referred to the Higher Education Quality and Standards division for any further action.

How to Avoid Plagiarism

Applying, analysing, criticising or quoting other people's work is perfectly reasonable and acceptable providing you always:

- attempt to summarize or restate in your own words another person's work, theories or ideas and give acknowledgement to that person. This is usually done by citing your sources and presenting a list of references;

or

- by always using quotation marks (or indenting lengthy quotations in your text) to distinguish between the actual words of the writer and your own words. Once again, you would cite all sources and present full details of these in your list of references.

What is a bibliography?

The term bibliography is the term used for a list of sources (e.g. books, articles, websites) used to write an assignment (e.g. an essay). It usually includes all the sources consulted even if they not directly cited (referred to) in the assignment.

How does a 'List of References' differ from a 'Bibliography'?

Your assignments will usually include a 'List of References' – this term is used for a list of sources that only includes those items you cite in your writing.

What about citing? What is that?

In your assignment, you will usually need to provide a citation in the text of your assignment when you are referring to evidence from one of your sources (references). This might be when you are discussing or summarising a theory or information in your own words, or you may be directly quoting from that source.

Avoid relying on lots of lengthy direct quotations Even if your in-text citations are accurate for direct quotations, these mostly show your reader you can identify a (relevant) quotation. Further discussion before and/or after the quote is required to make sure the significance and purpose of quotes is clear to your reader. How does the quote illustrate or support your argument?

Depending on the system you use, in-text citations may take the form of footnotes (at the bottom of the page of your writing), endnotes (that appear at the end of the written text of your assignment) or bracketed references (Jones, 2013).

Your in-text citation needs to give enough information so that your reader can then find the full details of the source of your evidence in your 'List of References' (or 'Bibliography'). The extent to which the information in your in-text citation will replicate the detail provided in the 'List of References' depends on the system of referencing. Systems using footnotes or endnotes commonly include more detail about the whole source than Harvard in-text references given in brackets. Harvard usually only requires the author's surname and date of publication, with page numbers provided if a "short direct quotation is included" (Jones, 2013, p.62).

Why does my feedback comment on my referencing being a problem?

It said my list of references at the end of the essay was fine. The word 'referencing' can be also be used instead of the term 'citing', so it is important to make sure that both your 'List of References' AND your in-text citations are appropriately presented. Common errors for in-text citations can be giving too much or too little information, depending on the referencing system you are supposed to be using.

How do I identify who the author is?

Usually, it is whoever is responsible for creating the content (writing the text). This may be an individual (or individuals), or it may be a collective or organisational group. Who is the publisher? The publisher is whoever/whatever is responsible for putting the item into the public domain, such as a book publisher (OUP; Routledge), a film or television studio (Miramax; Universal) or a corporate body (i.e. DoH – Department of Health).

How do I identify the place of publication?

Your 'List of References' / 'Bibliography' will usually require the place of publication as well as the name of the publisher for book printed items. The first listed town or city of publication is given rather than the country. Different referencing systems may recommend you also include additional information such as the US state. This can help distinguish between Cambridge, Massachusetts in the USA (Cambridge, Mass.) and Cambridge in the UK. The place of publication can usually be found on either:

- Frontispiece or title page: usually at the front of the book
- Copyright notice, publisher or printer's statement, or colophon (historically more common pre-1500): usually the verso (back) of the title page. N.B. Some textbooks publish this information on the final page at the back of the book

Self-Assessment Questions

1) Define research. How are research methods different from research methodology?
2) Discuss different approaches of research.
3) Differentiate between inductive and deductive; objective and subjective; quantitative and qualitative methods.
4) When deductive approach is used?
5) Discuss the different methods of sampling.
6) Make differences between reference and bibliography.
7) How is summary different from an abstract.
8) What is APA method of referencing?

Selected Readings

Ahuja, R. (2001). Research Methodology. Kolkata: Rawat Publication.

Bakshi, A. J. (2014). *Changing Paradigms in the Career Guidance Field: The Journey from Objectivity to Subjectivity, Quantitative to Qualitative Methods*. Indian Journal of Career and Livelihood Planning, 3(1), p. 32-41.

Burney, S. M. A. (2008). *Inductive and Deductive Research Approach,* accessed from http://www.drburney.net/INDUCTIVE%20&%20DEDUCTIVE%20RESEARCH%20APPROACH%2006032008.pdf.

Das, D. L. (2000). Practice of Social Research. New Delhi: Rawat Publication.

David, F. E. (2000). Scientific Method For Ecological Research. U.K: Cambridge.
Elder, S. (2009).*Sampling Methodology*. International Labour Organization, Geneva.

Etikan, I. & K. Bala (2017). *Sampling and Sampling Methods*. Biometrics & Biostatistics International Journal, 5(6).

Gelman, A. & C. Henning (2015). Beyond subjective and objective in Statistics.Available at http://stat.columbia.edu/~gelman/research/unpublished/objectivity10.pdf.

Gibaldi, J., & Achtert, W. S. (1989). MLA handbook for writers of research papers. New Delhi: Affiliated East West Press Private Limited.

Harper, C., & Marcus, R. (2007). Research for Development :A practical Guide . New Delhi: Vistaar Publication.

Khan, K. (2008).*Difference between Executive Summary, Abstract and Synopsis what Is Literature Review*, Accessed from https://www.scribd.com/doc/55954574/Difference-Between-Executive-Summary-Abstract-and-Synopsis on August 24, 2018.

Kothari, C.R. (1985). *Research Methodology- Methods and Techniques*, Wiley Eastern Limited, New Delhi.

Kumar, R. (2005). *Research Methodology-A Step-by-Step Guide for Beginners*, (2nd ed.), Pearson Education, Singapore.

Misra, R. (2001). Research Methodology:A handbook. New Delhi: Concept Publishing Company.

Mondal, R. Research Methodology for Social Scientist. Concept Publication.

Panneerselvam, R. (2009). Research Methodology. Learning private limited.

Raza, M. (1979). Survey of Research in Geography . Calcutta: Allied Publishers Private Limited.

Singh, K. (2007). Quantitative Social Research Methods . New Delhi: Sage Publication.

Somekh, B., & Lewin, C. (2005). Research Methods in the Social Science. New Delhi: Vistaar Publication.

Thomas, D. R. (2006).*A General Inductive Approach for Analyzing Qualitative Evaluation Data*. American Journal of Evaluation, 27(2), p. 237-46.

Webliography

1) https://www.diffen.com/difference/Objective_vs_Subjective
2) http://www.open.edu/openlearn/money-management/understanding-different-research-perspectives/content-section-1
3) https://keydifferences.com/difference-between-qualitative-and-quantitative-data.html
4) https://www.snapsurveys.com/blog/qualitative-vs-quantitative-research/
5) https://www.nottingham.ac.uk/studentservices/documents/whatarebibliographiesandreferences.pdf
6) https://grimsby.ac.uk/documents/quality/skills/DefinitiveGuideToHarvardReferencingAndBibliographies.pdf
7) https://in.sagepub.com/sites/default/files/upm-binaries/55588_Chapter_1_Sample_Creswell_Research_Design_4e.pdf
8) https://deborahgabriel.com/2013/03/17/inductive-and-deductive-approaches-to-research/
9) https://research-methodology.net/research-methodology/research-approach/inductive-approach-2/

10) https://research-methodology.net/research-methodology/research-approach/deductive-approach-2/

11) https://www.academia.edu/34646199/Importance_of_Inductive_and_Deductive_Approaches_in_Geographical_Researches

12) https://saylordotorg.github.io/text_principles-of-sociological-inquiry-qualitative-and-quantitative-methods/s05-03-inductive-or-deductive-two-dif.html

13) https://keydifferences.com/difference-between-objective-and-subjective.html

14) https://www.public.asu.edu/~kroel/www500/hypothesis.pdf

15) http://sola.siu.ac.th/sola/public/FacSites/GE1701/reading/Sampling.pdf

16) http://www.sjsu.edu/people/fred.prochaska/courses/ScWk240/s1/Week-6-Slides---Sampling.pdf

17) https://www.researchgate.net/publication/329179630_Identifying_and_Formulating_the_Research_Problem

18) https://www.adelaide.edu.au/writingcentre/sites/default/files/docs/learningguide-writinganabstract.pdf

19) https://depts.washington.edu/owrc/Handouts/How%20to%20Write%20a%20Summary.pdf

20) https://www.iue.edu/hss/writingcenter/documents/SummaryWriting.pdf

Disclaimer

This self-learning material is based on collections from different books, journals and web-sources.

Self-Learning Material (SLM) for the Course of

M.A. / M.Sc. in Geography

Directorate of Open and Distance Learning (DODL)

University of Kalyani

Paper: GEO311T(EG)

Special Paper Theory (Environmental Geography)

(Total Credit – 4; Total Marks – 100: Internal Evaluation – 20 + Semester-end Examination - 80)

University of Kalyani

Kalyani

Nadia – 741235

West Bengal

Course Material Compiled by

Dr. Bhaswati Mondal

Assistant Professor (Contractual) in Geography

Directorate of Open and Distance Learning

University of Kalyani

Kalyani

Nadia – 741235

Published by the Directorate of Open and Distance Learning (DODL), University of Kalyani, Kalyani – 741 235, West Bengal

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2017 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor (Dr.) Sankar Kumar Ghosh, Hon'ble Vice- Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks are also due to the course writers-cum-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and co-ordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self-writing and collected from e-book, journals and websites.

Director
Directorate of Open and Distance Learning
University of Kalyani

SYLLABUS

Paper: GEO311T(EG) – Special Paper Theory (Environmental Geography)

(Total Credit – 4; Total Marks – 100: Internal Evaluation – 20, Semester-end Examination - 80)

<u>Group – GEO311T (EG).1: CONCEPT</u>
(Credit – 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit-1: Definition of Relevant Terms: Ecosystem; Ecology

Unit-2: Nature, Scope and Content of Environmental Studies in Geography

Unit-3: Concept of Population Equilibrium, Optimum Population and Land-man Ratio

Unit-4: Gaia-hypothesis; Deep Ecology

Unit-5: Organismic and Holistic Explanations

Unit-6: Environmentalism in Geography; Spaceship Earth; Stationary State Economy

Unit-7: Concept of Environmental System

Unit-8: Environmental Balance and Environmental Degradation

<u>GROUP – GEO311T (EG).2: ENVIRONMENT AND DEVELOPMENT</u>

(Credit: 2; Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit-9: Man and Environment: Case Studies from River Valley Projects – Silent Valley and Narmada Dispute with Special Reference to Environmental Movement

Unit-10: Earth Summits: 1972, 1992, and 2012

Unit-11: Protocols: Montreal and Kyoto

Unit-12: Anthropogenic Impact on Environment: Population, Resource and Development

Unit-13: Environment; Environmental Impact Assessment; Environmental Audit and Environmental Management Planning

Unit-14: Concept and Methods of Alternative Agriculture

Unit-15: Use and Misuse of Forest Resources and Forest Conservation

Unit-16: Tourism Industry and Environment: Issues and Challenges

Contents

A. Introduction

The term environment etymologically means surrounding. Thus the entire natural features, social and cultural aspects which encircle us, are collectively known as environment. It is the particular branch of geography that describes and explains the spatial aspects of interactions between human individuals or societies and their natural environment. Environmental geography represents a critically important set of analytical tools for assessing the impact of human presence on the environment by measuring the result of human activity on natural landforms and cycles.

The links between human and physical geography were once more apparent than they are today. As human experience of the world is increasingly mediated by technology, the relationships between humans and the environment have often become obscured. Thereby, environmental geography represents a critically important set of analytical tools for assessing the impact of human presence on the environment. This is done by measuring the result of human activity on natural landforms and cycles.

B. Learning Objectives

After reading this special paper on Environmental Issues in Geography, the learners will be able to know in more or less details the concepts of

- Ecosystem and ecology,
- Nature, Scope and Content of Environmental Studies in Geography
- Population Equilibrium, Optimum Population and Land-man Ratio
- Gaia-hypothesis and Deep Ecology
- Organismic and Holistic Explanations
- Environmentalism in Geography
- Spaceship Earth
- Stationary State Economy
- Environmental System
- Environmental Balance
- Environmental Degradation
- River Valley Projects – Silent Valley and Narmada Dispute with Special Reference to Environmental Movement
- Earth Summits: 1972, 1992, and 2012
- Protocols: Montreal and Kyoto
- Anthropogenic Impact on Environment: Population, Resource and Development
- Environmental Impact Assessment
- Environmental Audit
- Environmental Management Planning
- Concept and Methods of Alternative Agriculture
- Use and Misuse of Forest Resources and Forest Conservation
- Tourism Industry and Environment: Issues and Challenges

C. Assessment of Prior Knowledge

To assess the students' prior knowledge, they may be asked

- What do you mean by environment?
- What are the major environmental issues at present?
- What are the requirements of studying environment?

D. Learning Activities

The learning process will involve Personal Contact Programmes, discussion, debate and interaction among students themselves, and students and teacher. During the Personal Contact Programmes, students may be assigned to prepare assignments on the issues of optimum population, deep ecology, criticism of the concept of deep ecology, comparative evaluation of different earth summits, environmental management planning, environmental audit etc.

E. Feedback of Learning Activities

Once the learning process is completed, internal assessments will be conducted. On the basis of evaluation reports of the internal assessments, some areas of the syllabus will be refocused depending on students' requirements.

Unit – 1

1.1 Definition of Relevant Terms: Ecosystem; Ecology

Introduction

Ecology is the scientific study of relationships in the natural world. It includes relationships between organisms and their physical environments (physiological ecology); between organisms of the same species (population ecology); between organisms of different species (community ecology); and between organisms and the fluxes of matter and energy through biological systems (ecosystem ecology). Ecologists study these interactions in order to understand the abundance and diversity of life within Earth's ecosystems—in other words, why there are so many plants and animals, and why there are so many different types of plants and animals.

To answer these questions they may use field measurements, such as counting and observing the behavior of species in their habitats; laboratory experiments that analyze processes such as predation rates in controlled settings; or field experiments, such as testing how plants grow in their natural setting but with different levels of light, water, and other inputs. Applied ecology uses information about these relationships to address issues such as developing effective vaccination strategies, managing fisheries without over-harvesting, designing land and marine conservation reserves for threatened species, and modeling how natural ecosystems may respond to global climate change.

Ecosystem

An ecosystem can be visualised as a functional unit of nature, where living organisms interact among themselves and also with the surrounding physical environment. Ecosystem varies greatly in size from a small pond to a large forest or a sea. Many ecologists regard the entire biosphere as a global ecosystem, as a composite of all local ecosystems on Earth.

Since this system is too much big and complex to be studied at one time, it is convenient to divide it into two basic categories, namely the terrestrial and the aquatic. Forest, grassland and desert are some examples of terrestrial ecosystems; pond, lake, wetland, river and estuary are some examples of aquatic ecosystems. Crop fields and an aquarium may also be considered as man-made ecosystems.

We will first look at the structure of the ecosystem, in order to appreciate the input (productivity), transfer of energy (food chain/web, nutrient cycling) and the output (degradation and energy loss). We will also look at the relationships – cycles, chains, webs – that are created as a result of these energy flows within the system and their inter- relationship.

Change is a constant process in ecosystems, driven by natural forces that include climate shifts, species movement, and ecological succession. By learning how ecosystems function, we can improve our ability to predict how they will respond to changes in the environment. But since living organisms in ecosystems are connected in complex relationships, it is not always easy to anticipate how a step such as introducing a new species will affect the rest of an ecosystem.

Components of Ecosystem

They are broadly grouped into:-

(a) Abiotic and

(b) Biotic components

(a) **Abiotic components (Non-living)**: The abiotic component can be grouped into following three categories:-

(i) Physical factors: Sun light, temperature, rainfall, humidity and pressure. They sustain and limit the growth of organisms in an ecosystem.

(ii) Inorganic substances: Carbon dioxide, nitrogen, oxygen, phosphorus, sulphur, water, rock, soil and other minerals.

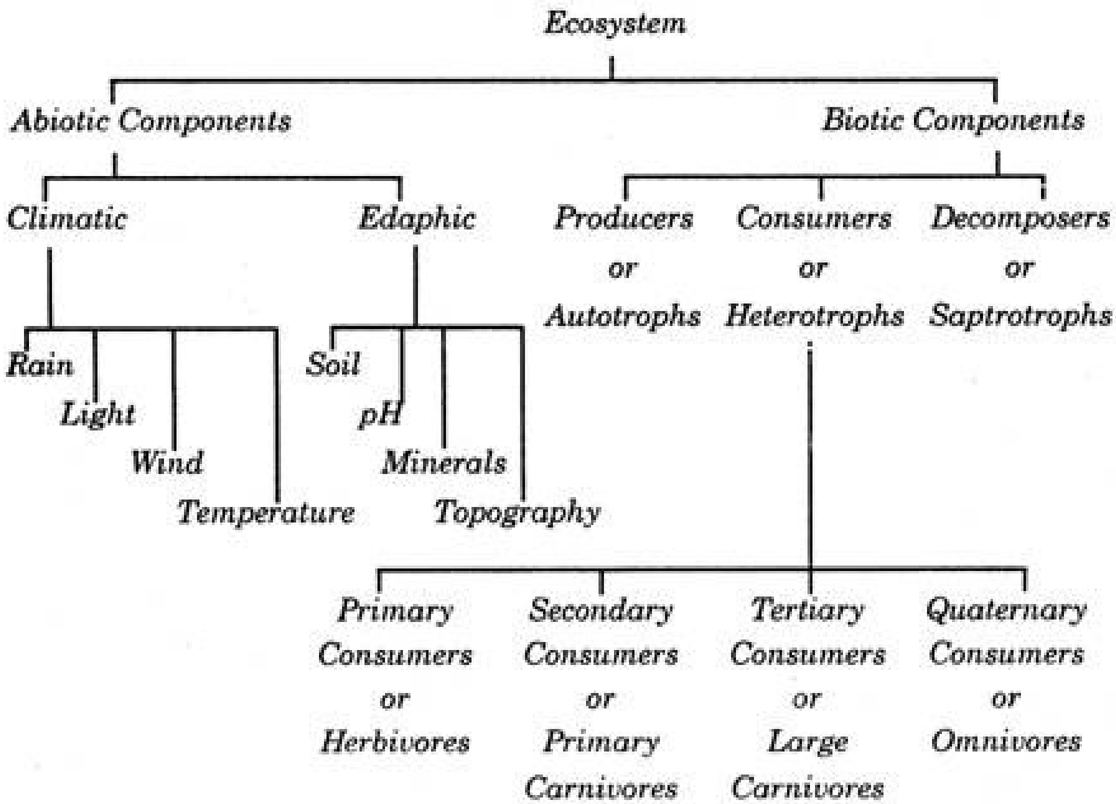
(iii) Organic compounds: Carbohydrates, proteins, lipids and humic substances. They are the building blocks of living systems and therefore, make a link between the biotic and abiotic components.

(b) **Biotic components (Living)**

(i) Producers: The green plants manufacture food for the entire ecosystem through the process of photosynthesis. Green plants are called autotrophs, as they absorb water and nutrients from the soil, carbon dioxide from the air, and capture solar energy for this process.

(ii) Consumers: They are called heterotrophs and they consume food synthesized by the autotrophs. Based on food preferences they can be grouped into three broad categories. Herbivores (e.g. cow, deer and rabbit etc.) feed directly on plants, carnivores are animals which eat other animals (eg. lion, cat, dog etc.) and omnivores organisms feeding upon both plants and animals e.g. human, pigs and sparrow etc.

(iii) Decomposers: Also called saprotrophs. These are mostly bacteria and fungi that feed on dead decomposed and the dead organic matter of plants and animals by secreting enzymes outside their body on the decaying matter. They play a very important role in recycling of nutrients. They are also called detrivores or detritus feeders.



Schematic Representation of the Structure of an Ecosystem.

Functions in Ecosystem

Interaction of biotic and abiotic components result in a physical structure that is characteristic for each type of ecosystem. Identification and enumeration of plant and animal species of an ecosystem gives its species composition. Vertical distribution of different species occupying different levels is called stratification. For example, trees occupy top vertical strata or layer of a forest, shrubs the second and herbs and grasses occupy the bottom layers. The components of the ecosystem are seen to function as a unit when you consider the following aspects:

(i) Productivity;

(ii) Decomposition;

(iii) Energy flow; and

(iv) Nutrient cycling.

To understand the ethos of an aquatic ecosystem let us take a small pond as an example. This is fairly a self-sustainable unit and rather simple example that explain even the complex interactions that exist in an aquatic ecosystem. A pond is a shallow water body in which all the above mentioned four basic components of an ecosystem are well exhibited.

The abiotic component is the water with all the dissolved inorganic and organic substances and the rich soil deposit at the bottom of the pond. The solar input, the cycle of temperature, day-length and other climatic conditions regulate the rate of function of the entire pond. The autotrophic components include the phytoplankton, some algae and the floating, submerged and marginal plants found at the edges. The consumers are represented by the zoo-plankton, the free swimming and bottom dwelling forms. The decomposers are the fungi, bacteria and flagellates especially abundant in the bottom of the pond.

This system performs all the functions of any ecosystem and of the biosphere as a whole, i.e., conversion of inorganic into organic material with the help of the radiant energy of the sun by the autotrophs; consumption of the autotrophs by heterotrophs; decomposition and mineralisation of the dead matter to release them back for reuse by the autotrophs, these event are repeated over and over again. There is unidirectional movement of energy towards the higher trophic levels and its dissipation and loss as heat to the environment.

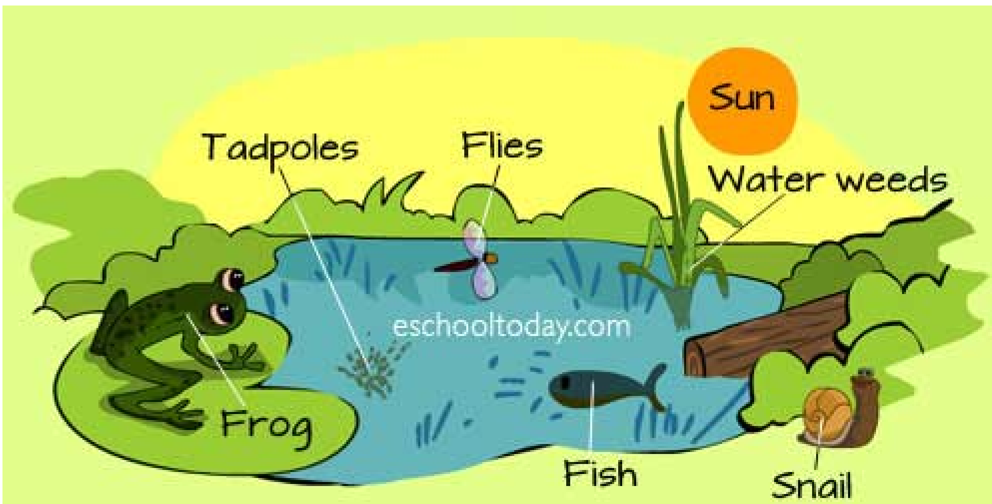


Figure: Ecosystem

PRODUCTIVITY

A constant input of solar energy is the basic requirement for any ecosystem to function and sustain. **Primary production** is defined as the amount of biomass or organic matter produced per unit area over a time period by plants during photosynthesis. It is expressed in terms of weight (g^{-2}) or energy (kcal m^{-2}). The rate of biomass production is called productivity. It is expressed in terms of g^{-2} yr^{-1} or (kcal m^{-2}) yr $^{-1}$ to compare the productivity of different ecosystems. It can be divided into gross primary productivity (GPP) and net primary productivity (NPP). Gross primary productivity of an ecosystem is the rate of production of organic matter during photosynthesis. A considerable amount of GPP is utilised by plants in respiration. Gross primary productivity minus respiration losses (R), is the net primary productivity (NPP). GPP – R = NPP Net primary productivity is the available biomass for the consumption to **heterotrophs** (herbiviores and decomposers).

Secondary productivity is defined as the rate of formation of new organic matter by consumers. Primary productivity depends on the plant species inhabiting a particular area. It also depends on a variety of environmental factors, availability of nutrients and photosynthetic capacity of plants. Therefore, it varies in different types of ecosystems. The annual net primary productivity of the whole biosphere is approximately 170 billion tons (dry weight) of organic matter. Of this, despite occupying about 70 per cent of the surface, the productivity of the oceans are only 55 billion tons. Rest of course, is on land. Discuss the main reason for the low productivity of ocean with your teacher.

DECOMPOSITION

You may have heard of the earthworm being referred to as the farmer's 'friend'. This is so because they help in the breakdown of complex organic matter as well as in loosening of the soil. Similarly, decomposers break down complex organic matter into inorganic substances like carbon dioxide, water and nutrients and the process is called decomposition. Dead plant remains such as leaves, bark, flowers and dead remains of animals, including fecal matter, constitute detritus, which is the raw material for decomposition. The important steps in the process of decomposition are fragmentation, leaching, catabolism, humification and mineralisation.

Detritivores (e.g., earthworm) break down detritus into smaller particles. This process is called fragmentation. By the process of leaching, water-soluble inorganic nutrients go down into the soil horizon and get precipitated as unavailable salts. Bacterial and fungal enzymes degrade detritus into simpler inorganic substances. This process is called as catabolism. It is important to note that all the above steps in decomposition operate simultaneously on the detritus. **Humification** and **mineralisation** occur during decomposition in the soil.

Humification leads to accumulation of a dark coloured amorphous substance called humus that is highly resistant to microbial action and undergoes decomposition at an extremely slow rate. Being colloidal in nature it serves as a reservoir of nutrients. The humus is further degraded by some microbes and release of inorganic nutrients occurs by the process known as **mineralisation**.

Decomposition is largely an oxygen-requiring process. The rate of decomposition is controlled by chemical composition of detritus and climatic factors. In a particular climatic condition, decomposition rate is slower if detritus is rich in lignin and chitin, and quicker, if detritus is rich in nitrogen and water-soluble substances like sugars. Temperature and soil moisture are the most important climatic factors that regulate decomposition through their effects on the activities of soil microbes. Warm and moist environment favour decomposition whereas low temperature and anaerobiosis inhibit decomposition resulting in build-up of organic materials.

ENERGY FLOW

Ecosystems maintain themselves by cycling energy and nutrients obtained from external sources. At the first trophic level, primary producers (plants, algae, and some bacteria) use solar energy to produce organic plant material through photosynthesis. Herbivores—animals that feed solely on plants—make up the second trophic level. Predators that eat herbivores comprise the third trophic level; if larger predators are present, they represent still higher trophic levels.

Organisms that feed at several trophic levels (for example, grizzly bears that eat berries and salmon) are classified at the highest of the trophic levels at which they feed. Decomposers, which include bacteria, fungi, molds, worms, and insects, break down wastes and dead organisms and return nutrients to the soil. On average about 10 per cent of net energy production at one trophic level is passed on to the next level. Processes that reduce the energy transferred between trophic levels include respiration, growth and reproduction, defecation, and non-predatory death (organisms that die but are not eaten by consumers). The nutritional quality of material that is consumed also influences how efficiently energy is transferred, because consumers can convert high-quality food sources into new living tissue more efficiently than low-quality food sources.

The low rate of energy transfer between trophic levels makes decomposers generally more important than producers in terms of energy flow. Decomposers process large amounts of organic material and return nutrients to the ecosystem in inorganic form, which is then taken up again by primary producers. Energy is not recycled during decomposition, but rather is released, mostly as heat (this is what makes compost piles and fresh garden mulch warm). Figure 6 shows the flow of energy (dark arrows) and nutrients (light arrows) through ecosystems.

Except for the deep sea hydro-thermal ecosystem, sun is the only source of energy for all ecosystems on Earth. Of the incident solar radiation less than 50 per cent of it is **photosynthetically active radiation** (PAR). We know that plants and photosynthetic bacteria (autotrophs), fix suns' radiant energy to make food from simple inorganic materials. Plants capture only 2-10 per cent of the PAR and this small amount of energy sustains the entire living world. So, it is very important to know how the solar energy captured by plants flows through different organisms of an ecosystem.

All organisms are dependent for their food on producers, either directly or indirectly. So you find unidirectional flow of energy from the sun to producers and then to consumers. Is this in keeping with the first law of thermodynamics? Further, ecosystems are not exempt from the Second Law of thermodynamics. They need a constant supply of energy to synthesise the molecules they require, to counteract the universal tendency toward increasing disorderliness. The green plant in the ecosystem-terminology are called **producers**.

In a terrestrial ecosystem, major producers are herbaceous and woody plants. Likewise, primary producers in an aquatic ecosystem are various species like phytoplankton, algae and higher plants. You have read about the **food chains** and **webs** that exist in nature. Starting from the plants (or producers) food chains or rather webs are formed such that an animal feeds on a plant or on another animal and in turn is food for another. The chain or web is formed because of this interdependency. No energy that is trapped into an organism remains in it for ever.

The energy trapped by the producer, hence, is either passed on to a consumer or the organism dies. Death of organism is the beginning of the detritus food chain/web. All animals depend on plants (directly or indirectly) for their food needs. They are hence called consumers and also heterotrophs. If they feed on the producers, the plants, they are called primary consumers, and if the animals eat other animals which in turn eat the plants (or their produce) they are called secondary consumers. Likewise, you could have tertiary consumers too.

Obviously the primary consumers will be herbivores. Some common herbivores are insects, birds and mammals in terrestrial ecosystem and molluscs in aquatic ecosystem. The consumers that feed on these herbivores are carnivores, or more correctly primary carnivores (though secondary consumers). Those animals that depend on the primary carnivores for food are labelled secondary carnivores.

The **detritus food chain (DFC**) begins with dead organic matter. It is made up of decomposers which are heterotrophic organisms, mainly fungi and bacteria. They meet their energy and nutrient requirements by degrading dead organic matter or detritus. These are also known as saprotrophs (sapro: to decompose).

Decomposers secrete digestive enzymes that breakdown dead and waste materials into simple, inorganic materials, which are subsequently absorbed by them. In an aquatic ecosystem, GFC is the major conduit for energy flow. As against this, in a terrestrial ecosystem, a much larger fraction of energy flows through the detritus food chain than through the GFC. Detritus food chain may be connected with the grazing food chain at some levels: some of the organisms of DFC are prey to the GFC animals, and in a natural ecosystem, some animals like cockroaches, crows, etc., are omnivores.

These natural interconnection of food chains make it a food web. How would you classify human beings! Organisms occupy a place in the natural surroundings or in a community according to their feeding relationship with other organisms. Based on the source of their nutrition or food, organisms occupy a specific place in the food chain that is known as their trophic level. Producers belong to the first trophic level, herbivores (primary consumer) to the second and carnivores (secondary consumer) to the third.

The important point to note is that the amount of energy decreases at successive trophic levels. When any organism dies it is converted to detritus or dead biomass that serves as an energy source for decomposers. Organisms at each trophic level depend on those at the lower trophic level for their energy demands. Each trophic level has a certain mass of living material at a particular time called as the standing crop.

The standing crop is measured as the mass of living organisms (biomass) or the number in a unit area. The biomass of a species is expressed in terms of fresh or dry weight. Measurement of biomass in terms of dry weight is more accurate. Why? The number of trophic levels in the grazing food chain is restricted as the transfer of energy follows 10 per cent law – only 10 per cent of the energy is transferred to each trophic level from the lower trophic level. In nature, it is possible to have so many levels – producer, herbivore, primary carnivore, secondary carnivore in the grazing food chain. Do you think there is any such limitation in a detritus food chain?

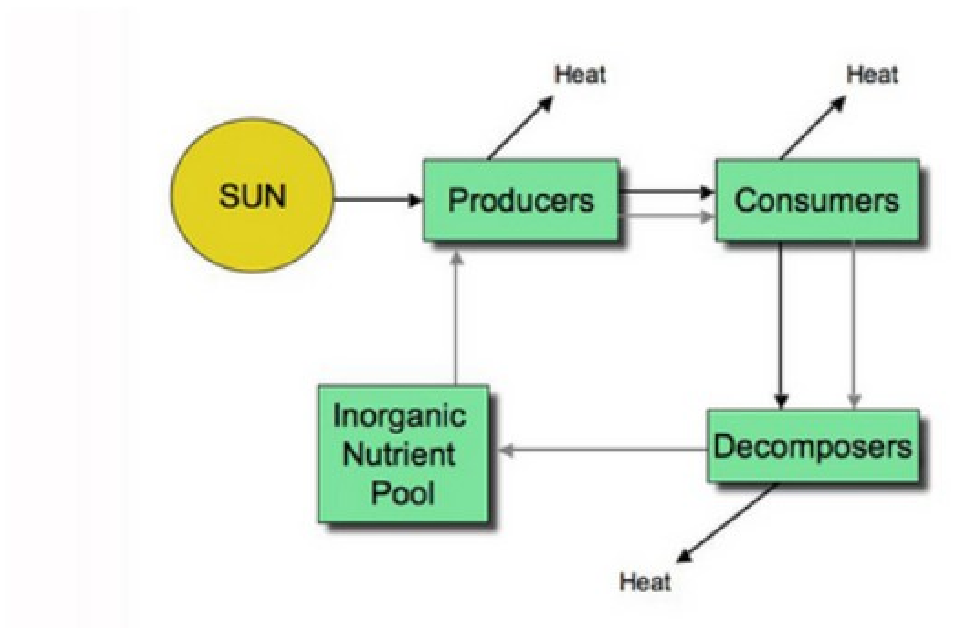


Figure 6. Energy and nutrient transfer through ecosystems

© Ohio Environmental Protection Agency. Nature Connections.

An ecosystem's gross primary productivity (GPP) is the total amount of organic matter that it produces through photosynthesis. Net primary productivity (NPP) describes the amount of energy that remains available for plant growth after subtracting the fraction that plants use for respiration.

Productivity in land ecosystems generally rises with temperature up to about 30°C, after which it declines, and is positively correlated with moisture. On land primary productivity thus is highest in warm, wet zones in the tropics where tropical forest biomes are located. In contrast, desert scrub ecosystems have the lowest productivity because their climates are extremely hot and dry.

In the oceans, light and nutrients are important controlling factors for productivity. "Oceans," light penetrates only into the uppermost level of the oceans, so photosynthesis occurs in surface and near-surface waters.

Marine primary productivity is high near coastlines and other areas where upwelling brings nutrients to the surface, promoting plankton blooms. Runoff from land is also a source of nutrients in estuaries and along the continental shelves. Among aquatic ecosystems, algal beds and coral reefs have the highest net primary production, while the lowest rates occur in the open due to a lack of nutrients in the illuminated surface layers How many trophic levels can an ecosystem support? The answer depends on several factors, including the amount of energy entering the ecosystem, energy loss between trophic levels, and the form, structure, and physiology of organisms at each level.

At higher trophic levels, predators generally are physically larger and are able to utilize a fraction of the energy that was produced at the level beneath them, so they have to forage over increasingly large areas to meet their caloric needs. Because of these energy losses, most terrestrial ecosystems have no more than five trophic levels, and marine ecosystems generally have no more than seven. This difference between terrestrial and marine ecosystems is likely due to differences in the fundamental characteristics of land and marine primary organisms.

In marine ecosystems, microscopic phytoplankton carry out most of the photosynthesis that occurs, while plants do most of this work on land. Phytoplankton are small organisms with extremely simple structures, so most of their primary production is consumed and used for energy by grazing organisms that feed on them. In contrast, a large fraction of the biomass that land plants produce, such as roots, trunks, and branches, cannot be used by herbivores for food, so proportionately less of the energy fixed through primary production travels up the food chain.

Growth rates may also be a factor. Phytoplankton are extremely small but grow very rapidly, so they support large populations of herbivores even though there may be fewer algae than herbivores at any given moment. In contrast, land plants may take years to reach maturity, so an average carbon atom spends a longer residence time at the primary producer level on land than it does in a marine ecosystem. In addition, locomotion costs are generally higher for terrestrial organisms compared to those in aquatic environments. The simplest way to describe the flux of energy through ecosystems is as a food chain in which energy passes from one trophic level to the next, without factoring in more complex relationships between individual species.

Some very simple ecosystems may consist of a food chain with only a few trophic levels. For example, the ecosystem of the remote wind-swept Taylor Valley in Antarctica consists mainly of bacteria and algae that are eaten by nematode worms. More commonly, however, producers and consumers are connected in intricate food webs with some consumers feeding at several trophic levels An important consequence of the loss of energy between trophic levels is that contaminants collect in animal tissues—a process called bioaccumulation. As contaminants bioaccumulate up the food web, organisms at higher trophic levels can be threatened even if the pollutant is introduced to the environment in very small quantities. The insecticide DDT, which was widely used in the United States from the 1940s through the 1960s, is a famous case of bioaccumulation.

DDT built up in eagles and other raptors to levels high enough to affect their reproduction, causing the birds to lay thin-shelled eggs that broke in their nests. Fortunately, populations have rebounded over several decades since the pesticide was banned in the United States. However, problems persist in some developing countries where toxic **bio-accumulating pesticides** are still used. Bioaccumulation can threaten humans as well as animals. For example, in the United States many federal and state agencies currently warn consumers to avoid or limit their consumption of large predatory fish that contain high levels of mercury, such as shark, swordfish, tilefish, and king mackerel, to avoid risking neurological damage and birth defects.

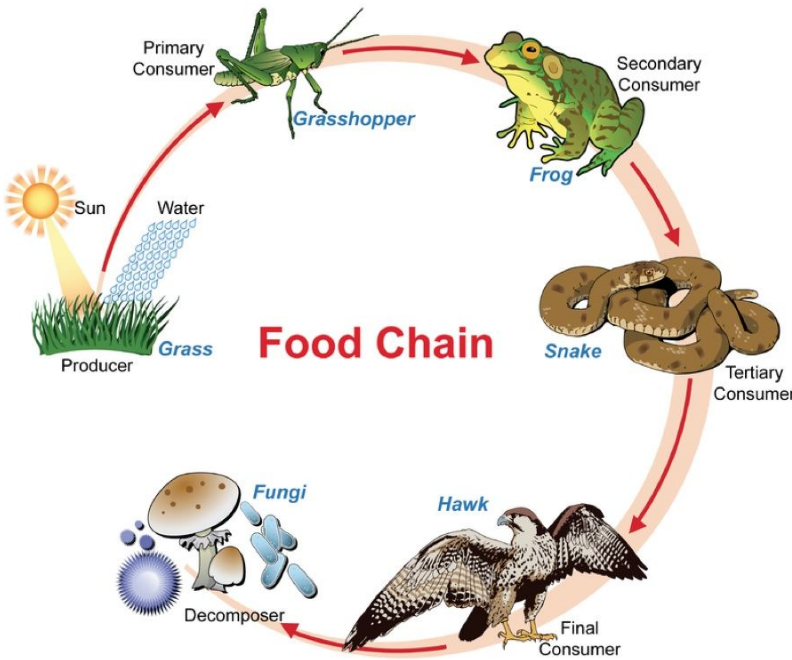


Figure: Food chain in a grassland ecosystem

Biogeochemical Cycling in Ecosystems

Along with energy, water and several other chemical elements cycle through ecosystems and influence the rates at which organisms grow and reproduce. About 10 major nutrients and six trace nutrients are essential to all animals and plants, while others play important roles for selected species.

The most important biogeochemical cycles affecting ecosystem health are the water, carbon, nitrogen, and phosphorus cycles. As noted earlier, most of the Earth's area that is covered by water is ocean. In terms of volume, the oceans dominate further still: nearly all of Earth's water inventory is contained in the oceans (about 97 percent) or in ice caps and glaciers (about 2 percent), with the rest divided among groundwater, lakes, rivers, streams, soils, and the atmosphere. In addition, water moves very quickly through land ecosystems.

These two factors mean that water's residence time in land ecosystems is generally short, on average one or two months as soil moisture, weeks or months in shallow groundwater, or up to six months as snow cover. But land ecosystems process a lot of water: almost two-thirds of the water that falls on land as precipitation annually is transpired back into the atmosphere by plants, with the rest flowing into rivers and then to the oceans.

Because cycling of water is central to the functioning of land ecosystems, changes that affect the hydrologic cycle are likely to have significant impacts on land ecosystems. (Global water cycling is discussed in more detail in Unit 8, "Water Resources.") Both land and ocean ecosystems are important sinks for carbon, which is taken up by plants and algae during photosynthesis and fixed as plant tissue. Table 2 compares the quantities of carbon stored in Earth's major reservoirs.

ECOLOGICAL PYRAMIDS

You must be familiar with the shape of a pyramid. The base of a pyramid is broad and it narrows down at the apex. Thus, relationship is expressed in terms of number, biomass or energy. The base of each pyramid represents the producers or the first trophic level while the apex represents tertiary or top level consumer. The three ecological pyramids that are usually studied are

(a) pyramid of number;

(b) pyramid of biomass and

(c) pyramid of energy.

Any calculations of energy content, biomass, or numbers has to include all organisms at that trophic level. No generalisations we make will be true if we take only a few individuals at any trophic level into account. Also a given organism may occupy more than one trophic level simultaneously. One must remember that the trophic level represents a functional level, not a species as such.

A given species may occupy more than one trophic level in the same ecosystem at the same time; for example, a sparrow is a primary consumer when it eats seeds, fruits, peas, and a secondary consumer when it eats insects and worms. Can you work out how many trophic levels human beings function at in a food chain? In most ecosystems, all the pyramids, of number, of energy and biomass are upright, i.e., producers are more in number and biomass than the herbivores, and herbivores are more in number and biomass than the carnivores.

Also energy at a lower trophic level is always more than at a higher level. There are exceptions to this generalisation: If you were to count the number of insects feeding on a big tree what kind of pyramid would you get? Now add an estimate of the number of small birds depending on the insects, as also the number of larger birds eating the smaller. Draw the shape you would get.

The pyramid of biomass in sea is also generally inverted because the biomass of fishes far exceeds that of phytoplankton. Isn't that a paradox? How would you explain this? Pyramid of energy is always upright, can never be inverted, because when energy flows from a particular trophic level to the next trophic level, some energy is always lost as heat at each step. Each bar in the energy pyramid indicates the amount of energy present at each trophic level in a given time or annually per unit area.

However, there are certain limitations of ecological pyramids such as it does not take into account the same species belonging to two or more trophic levels. It assumes a simple food chain, something that almost never exists in nature; it does not accommodate a food web. Moreover, saprophytes are not given any place in ecological pyramids even though they play a vital role in the ecosystem.

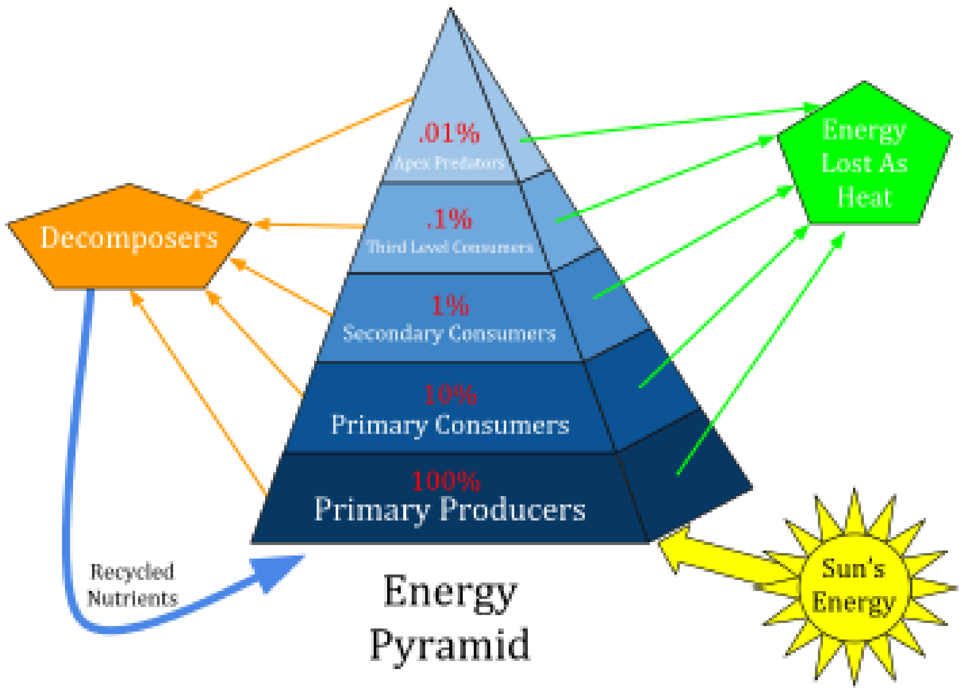


Figure: Energy Pyramid in an Ecosystem

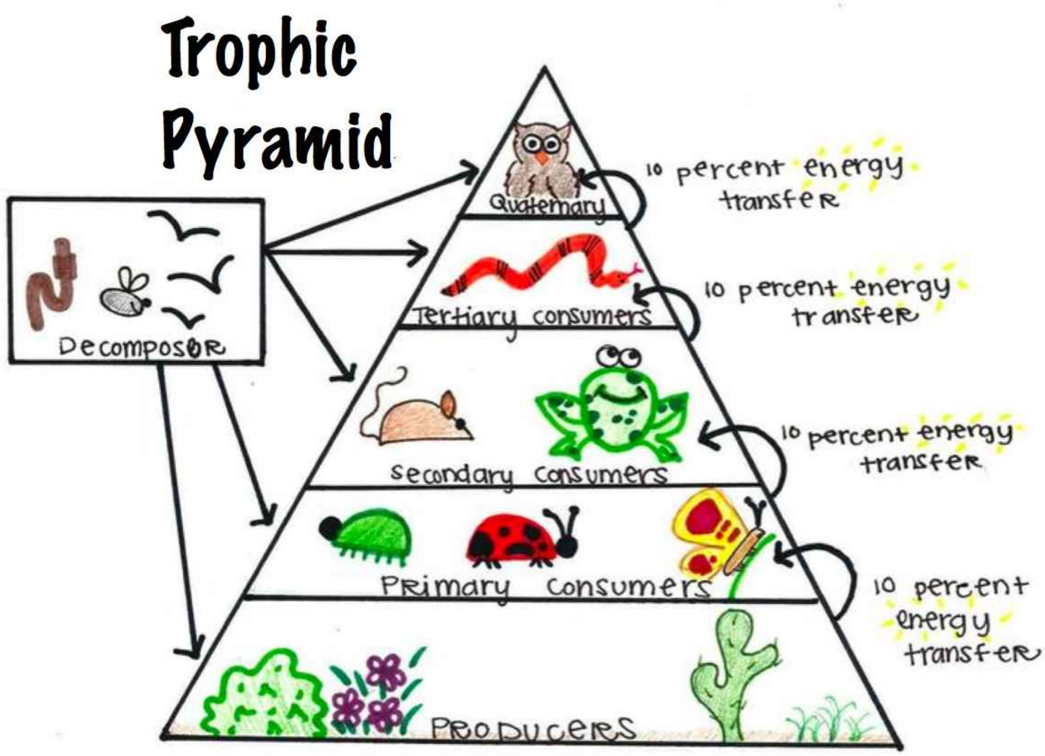
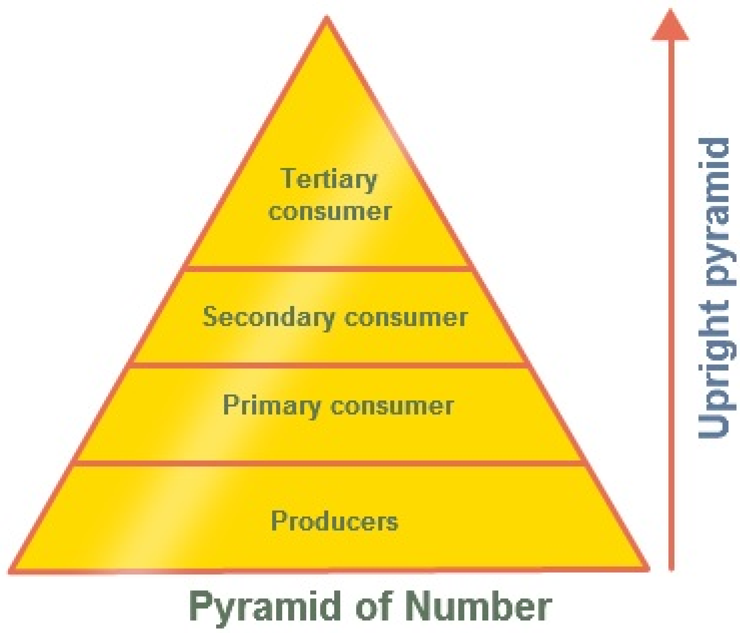


Figure: Tropihic Pyramid



ECOLOGICAL SUCCESSION

An important characteristic of all communities is that their composition and structure constantly change in response to the changing environmental conditions. This change is orderly and sequential, parallel with the changes in the physical environment. These changes lead finally to a community that is in near equilibrium with the environment and that is called a climax community. The gradual and fairly predictable change in the species composition of a given area is called ecological succession. During succession some species colonise an area and their populations become more numerous, whereas populations of other species decline and even disappear. The entire sequence of communities that successively change in a given area are called sere(s).

The individual transitional communities are termed seral stages or seral communities. In the successive seral stages there is a change in the diversity of species of organisms, increase in the number of species and organisms as well as an increase in the total biomass. The present day communities in the world have come to be because of succession that has occurred over millions of years since life started on earth. Actually succession and evolution would have been parallel processes at that time. Succession is hence a process that starts where no living organisms are there – these could be areas where no living organisms ever existed, say bare rock; or in areas that somehow, lost all the living organisms that existed there.

The former is called primary succession, while the latter is termed secondary succession. Examples of areas where primary succession occurs are newly cooled lava, bare rock, newly created pond or reservoir. The establishment of a new biotic community is generally slow. Before a biotic community of diverse organisms can become established, there must be soil. Depending mostly on the climate, it takes natural processes several hundred to several thousand years to produce fertile soil on bare rock.

Secondary succession begins in areas where natural biotic communities have been destroyed such as in abandoned farm lands, burned or cut forests, lands that have been flooded. Since some soil or sediment is present, succession is faster than primary succession. Description of ecological succession usually focuses on changes in vegetation. However, these vegetational changes in turn affect food and shelter for various types of animals.

Thus, as succession proceeds, the numbers and types of animals and decomposers also change. At any time during primary or secondary succession, natural or human induced disturbances (fire, deforestation, etc.), can convert a particular seral stage of succession to an earlier stage. Also such disturbances create new conditions that encourage some species and discourage or eliminate other species.

Succession of Plants

Based on the nature of the habitat – whether it is water (or very wet areas) or it is on very dry areas – succession of plants is called hydrarch or xerarch, respectively.

Hydrarch succession takes place in wetter areas and the successional series progress from hydric to the mesic conditions.

As against this, xerarch succession takes place in dry areas and the series progress from xeric to mesic conditions. Hence, both hydrarch and xerarch successions lead to medium water conditions (mesic) – neither too dry (xeric) nor too wet (hydric).

The species that invade a bare area are called pioneer species. In primary succession on rocks these are usually lichens which are able to secrete acids to dissolve rock, helping in weathering and soil formation. These later pave way to some very small plants like bryophytes, which are able to take hold in the small amount of soil. They are, with time, succeeded by bigger plants, and after several more stages, ultimately a stable climax forest community is formed.

The **climax community** remains stable as long as the environment remains unchanged. With time the xerophytic habitat gets converted into a mesophytic one. In primary succession in water, the pioneers are the small phytoplanktons, they are replaced with time by rooted-submerged plants, rooted-floating angiosperms followed by free-floating plants, then reedswamp, marsh-meadow, scrub and finally the trees. The climax again would be a forest. With time the water body is converted into land.

In secondary succession the species that invade depend on the condition of the soil, availability of water, the environment as also the seeds or other propagules present. Since soil is already there, the rate of succession is much faster and hence, climax is also reached more quickly. What is important to understand is that succession, particularly primary succession, is a very slow process, taking maybe thousands of years for the climax to be reached. Another important fact is to understand that all succession whether taking place in water or on land, proceeds to a similar climax community – the mesic.

NUTRIENT CYCLING

The amount of nutrients, such as carbon, nitrogen, phosphorus, calcium, etc., present in the soil at any given time, is referred to as the standing state. It varies in different kinds of ecosystems and also on a seasonal basis. What is important is to appreciate that nutrients which are never lost from the ecosystems, are recycled time and again indefinitely.

The movement of nutrient elements through the various components of an ecosystem is called nutrient cycling. Another name of nutrient cycling is biogeochemical cycles (bio: living organism, geo: rocks, air, water). Nutrient cycles are of two types: (a) gaseous and (b) sedimentary. The reservoir for gaseous type of nutrient cycle (e.g., nitrogen, carbon cycle) exists in the atmosphere and for the sedimentary cycle (e.g., sulphur and phosphorus cycle), the reservoir is located in Earth's crust. Environmental factors, e.g., soil, moisture, pH, temperature, etc., regulate the rate of release of nutrients into the atmosphere. The function of the reservoir is to meet with the deficit which occurs due to imbalance in the rate of influx and efflux. You have made a detailed study of nitrogen cycle in class XI. Here we discuss carbon and phosphorus cycles.

- **Ecosystem – Carbon Cycle**

When you study the composition of living organisms, carbon constitutes 49 per cent of dry weight of organisms and is next only to water. If we look at the total quantity of global carbon, we find that 71 per cent carbon is found dissolved in oceans. This oceanic reservoir regulates the amount of carbon dioxide in the atmosphere.

Do you know that the atmosphere only contains about 1per cent of total global carbon? Fossil fuel also represent a reservoir of carbon. Carbon cycling occurs through atmosphere, ocean and through living and dead organisms. According to one estimate 4 × 1013 kg of carbon is fixed in the biosphere through photosynthesis annually.

A considerable amount of carbon returns to the atmosphere as CO2 through respiratory activities of the producers and consumers. Decomposers also contribute substantially to CO2 pool by their processing of waste materials and dead organic matter of land or oceans. Some amount of the fixed carbon is lost to sediments and removed from circulation. Burning of wood, forest fire and combustion of organic matter, fossil fuel, volcanic activity are additional sources for releasing CO2 in the atmosphere. Human activities have significantly influenced the carbon cycle. Rapid deforestation and massive burning of fossil fuel for energy and transport have significantly increased the rate of release of carbon dioxide into the atmosphere.

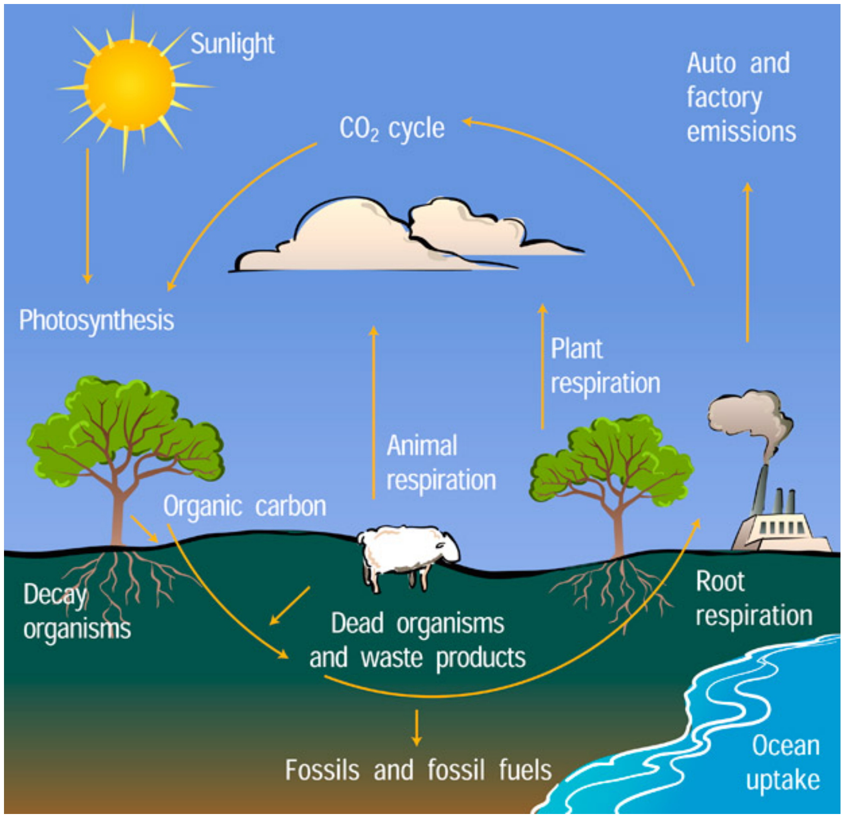


Figure: Carbon Cycling

- **Ecosystem – Phosphorus Cycle**

Phosphorus is a major constituent of biological membranes, nucleic acids and cellular energy transfer systems. Many animals also need large quantities of this element to make shells, bones and teeth. The natural reservoir of phosphorus is rock, which contains phosphorus in the form of phosphates. When rocks are weathered, minute amounts of these phosphates dissolve in soil solution and are absorbed by the roots of the plants.

Herbivores and other animals obtain this element from plants. The waste products and the dead organisms are decomposed by phosphate-solubilising bacteria releasing phosphorus. Unlike carbon cycle, there is no respiratory release of phosphorus into atmosphere. Can you differentiate between the carbon and the phosphorus cycle? The other two major and important differences between carbon and phosphorus cycle are firstly, atmospheric inputs of phosphorus through rainfall are much smaller than carbon inputs, and, secondly, gaseous exchanges of phosphorus between organism and environment are negligible.

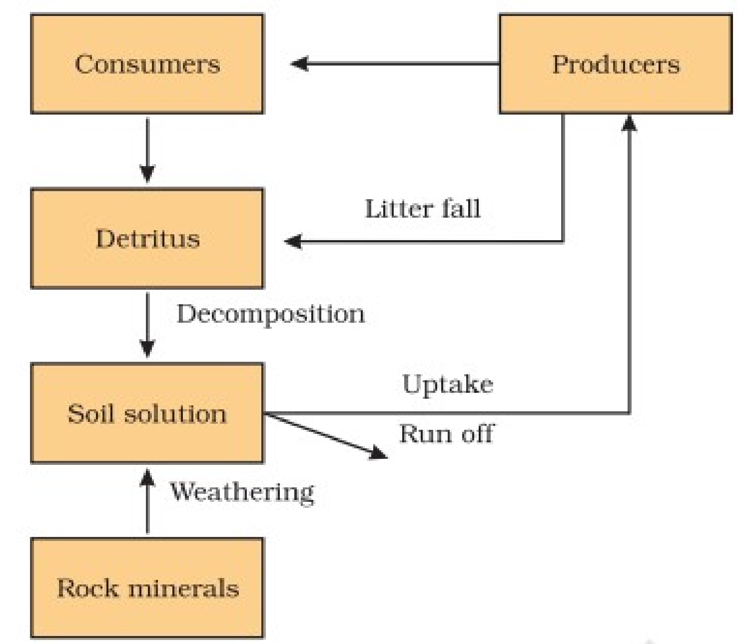


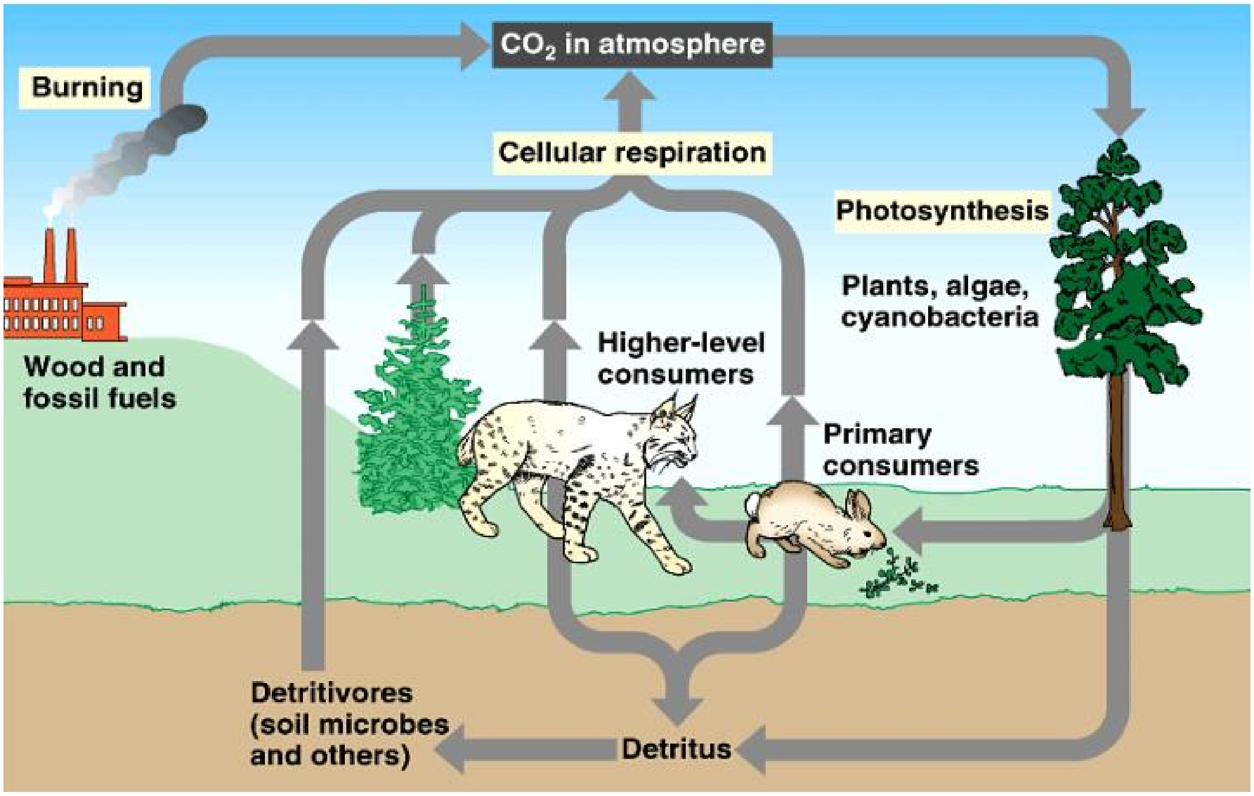
Figure: Phosphorus Cycling in Terrestrial Ecosystem

Ecosystem Dynamics

When new biotic or abiotic elements enter into an ecosystem, they cause an interruption. This can also lead to death of certain species within the ecosystem. But often ecosystems can protect themselves from intruders, depending on the toxicity of the new element and the resiliency of the original ecosystem.

Organisms have survived despite continuous changes, natural selection and intruders, but they had to adapt to new conditions. Changes are always gradual. Some species even disappear and new ones move in. Usually the population remains within limits of the food supply within an ecosystem. In general, ecosystems are more resilient to sudden or great changes than each species individually. Arctic tundra in Russia is an example of an ecosystem which has remained relatively unchanged for a long period of time.

Ecosystem Dynamics



Population Dynamics

Every organism in an ecosystem divides its energy among three competing goals: growing, surviving, and reproducing. Ecologists refer to an organism's allocation of energy among these three ends throughout its lifetime as its life history strategy. There are trade-offs between these functions: for example, an organism that spends much of its energy on reproduction early in life will have lower growth and survival rates, and thus a lower reproductive level later in life. An optimal life history strategy maximizes the organism's contribution to population growth. Understanding how the environment shapes organisms' life histories is a major question in ecology. Compare the conditions for survival in an unstable area, such as a flood plain near a river that frequently overflows its banks, to those in a stable environment, such as a remote old-growth forest. On the flood plain, there is a higher chance of being killed early in life, so the organisms that mature and reproduce earlier will be most likely to survive and add to population growth. Producing many offspring increases the chance that some will survive. Conversely, organisms in the forest will mature later and have lower early reproductive rates. This allows them to put more energy into growth and competition for resources.

Ecologists refer to organisms at the first of these two extremes (those adapted to unstable environments) as r-selected. These organisms live in settings where population levels are well below the maximum number that the environment can support—the carrying capacity—so their numbers are growing exponentially at the maximum rate at which that population can increase if resources are not limited (often abbreviated as r). The other extreme, organisms adapted to stable environments, are termed

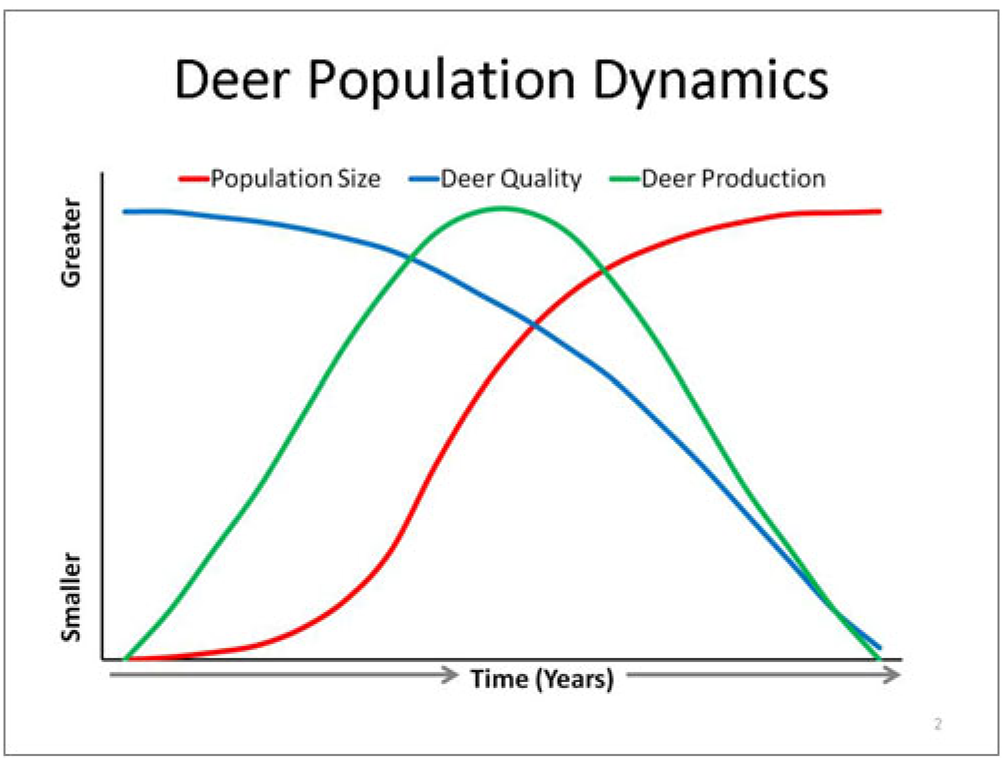
K-selected because they live in environments in which the number of individuals is at or near the environment's carrying capacity (often abbreviated as K).

Organisms that are r-selected tend to be small, short-lived, and opportunistic, and to grow through irregular boom-and-bust population cycles. They include many insects, annual plants, bacteria, and larger species such as frogs and rats. Species considered pests typically are r-selected organisms that are capable of rapid growth when environmental conditions are favorable.

In contrast, K-selected species are typically larger, grow more slowly, have fewer offspring and spend more time parenting them. Examples include large mammals, birds, and long-lived plants such as redwood trees. Kselected species are more prone to extinction than r-selected species because they mature later in life and have fewer offspring with longer gestation times. Table 3 contrasts the reproductive characteristics of an r-selected mammal, the Norway rat, to those of a K-selected mammal, the African elephant.

Many organisms fall between these two extremes and have some characteristics of both types. As we will see below, ecosystems tend to be dominated by r-selected species in their early stages with the balance gradually shifting toward K-selected species. In a growing population, survival and reproduction rates will not stay constant over time.

Eventually resource limitations will reduce one or both of these variables. Populations grow fastest when they are near zero and the species is uncrowded. A simple mathematical model of population growth implies that the maximum population growth rate occurs when the population size (N) is at one-half of the environment's carrying capacity, K (i.e., at $N = K/2$). In theory, if a population is harvested at exactly its natural rate of growth, the population will not change in size, and the harvest (yield) can be sustained at that level. In practice, however, it can be very hard to estimate population sizes and growth rates in the wild accurately enough to achieve this maximum sustainable yield.



Regulation of Ecosystem Functions

A key question for ecologists studying growth and productivity in ecosystems is which factors limit ecosystem activity. Availability of resources, such as light, water, and nutrients, is a key control on growth and reproduction. Some nutrients are used in specific ratios. For example, the ratio of nitrogen to phosphorus in the organic tissues of algae is about 16 to 1, so if the available nitrogen concentration is greater than 16 times the phosphorus concentration, then phosphorus will be the factor that limits growth; if it is less, then nitrogen will be limiting.

To understand how a specific ecosystem functions, it thus is important to identify what factors limit ecosystem activity. Resources influence ecosystem activity differently depending on whether they are essential, substitutable, or complementary. Essential resources limit growth independently of other levels: if the minimum quantity needed for growth is not available, then growth does not occur. In contrast, if two resources are substitutable, then population growth is limited by an appropriately weighted sum of the two resources in the environment.

For example, glucose and fructose are substitutable food sources for many types of bacteria. Resources may also be complementary, which means that a small amount of one resource can substitute for a relatively large amount of another, or can be complementary over a specific range of conditions.

Resource availability serves as a so-called "bottom-up" control on an ecosystem: the supply of energy and nutrients influences ecosystem activities at higher trophic levels by affecting the amount of energy that moves up the food chain. In some cases, ecosystems may be more strongly influenced by so-called "top-down" controls—namely, the abundance of organisms at high trophic levels in the ecosystem. Both types of effects can be at work in an ecosystem at the same time, but how far bottom-up effects extend in the food web, and the extent to which the effects of trophic interactions at the top of the food web are felt through lower levels, vary over space and time and with the structure of the ecosystem.

Many ecological studies seek to measure whether bottom-up or top-down controls are more important in specific ecosystems because the answers can influence conservation and environmental protection strategies. For example, a study by Benjamin S. Halpern and others of food web controls in kelp forest ecosystems off the coast of Southern California found that variations in predator abundance explained a significant proportion of variations in the abundance of algae and the organisms at higher trophic levels that fed on algae and plankton. In contrast, they found no significant relationship between primary production by algae and species abundance at higher trophic levels.

The most influential predators included spiny lobster, Kellet's whelk, rockfish, and sea perch. Based on these findings, the authors concluded that "[e]fforts to control activities that affect higher trophic levels (such as fishing) will have far larger impacts on community dynamics than efforts to control, for example, nutrient input, except when these inputs are so great as to create anoxic (dead) zones" (footnote 4). Drastic changes at the top of the food web can trigger trophic cascades, or domino effects that are felt through many lower trophic levels.

The likelihood of a trophic cascade depends on the number of trophic levels in the ecosystem and the extent to which predators reduce the abundance of a trophic level to below their resource-limited carrying capacity. Some species are so important to an entire ecosystem that they are referred to as keystone species, connoting that they occupy an ecological niche that influences many other species. Removing or seriously impacting a keystone species produces major impacts throughout the ecosystem. Many scientists believe that the reintroduction of wolves into Yellowstone National Park in 1995, after they had been eradicated from the park for decades through hunting, has caused a trophic cascade with results that are generally positive for the ecosystem.

Wolves have sharply reduced the population of elk, allowing willows to grow back in many riparian areas where the elk had grazed the willows heavily. Healthier willows are attracting birds and small mammals in large numbers. "Species, like riparian songbirds, insects, and in particular, rodents, have come back into these preferred habitat types, and other species are starting to respond," says biologist Robert Crabtree of the Yellowstone Ecological Research Center. "For example, fox and coyotes are moving into these areas because there's more prey for them. There's been an erupting trophic cascade in some of these lush riparian habitat sites."

ECOSYSTEM SERVICES

Healthy ecosystems are the base for a wide range of economic, environmental and aesthetic goods and services. The products of ecosystem processes are named as ecosystem services, for example, healthy forest ecosystems purify air and water, mitigate droughts and floods, cycle nutrients, generate fertile soils, provide wildlife habitat, maintain biodiversity, pollinate crops, provide storage site for carbon and also provide aesthetic, cultural and spiritual values. Though value of such services of biodiversity is difficult to determine, it seems reasonable to think that biodiversity should carry a hefty price tag. Robert Constanza and his colleagues have very recently tried to put price tags on nature's life-support services. Researchers have put an average price tag of US $ 33 trillion a year on these fundamental ecosystems services, which is largely taken for granted because they are free.

This is nearly twice the value of the global gross national product GNP which is (US $ 18 trillion). Out of the total cost of various ecosystem services, the soil formation accounts for about 50 per cent, and contributions of other services like recreation and nutrient cycling, are less than 10 per cent each. The cost of climate regulation and habitat for wildlife are about 6 per cent each.

Figure: Ecosystem Services

Ecological Niches

Each species in an ecosystem occupies a niche, which comprises the sum total of its relationships with the biotic and abiotic elements of its environment—more simply, what it needs to survive. In a 1957 address, zoologist George Evelyn Hutchinson framed the view that most ecologists use today when he defined the niche as the intersection of all of the ranges of tolerance under which an organism can live.

This approach makes ecological niches easier to quantify and analyze because they can be described as specific ranges of variables like temperature, latitude, and altitude. For example, the African Fish Eagle occupies a very similar ecological niche to the American Bald Eagle. In practice it is hard to measure all of the variables that a species needs to survive, so descriptions of an organism's niche tend to focus on the most important limiting factors.

The full range of habitat types in which a species can exist and reproduce without any competition from other species is called its fundamental niche. The presence of other species means that few species live in such conditions. A species' realized niche can be thought of as its niche in practice — the range of habitat types from which it is not excluded by competing species. Realized niches are usually smaller than fundamental niches, since competitive interactions exclude species from at least some conditions under which they would otherwise grow. Species may occupy different realized niches in various locations if some constraint, such as a certain predator, is present in one area but not in another. In a classic set of laboratory experiments, Russian biologist G.F. Gause showed the difference between fundamental and realized niches.

Gause compared how two strains of Paramecium grew when they were cultured separately in the same type of medium to their growth rates when cultured together. When cultured separately both strains reproduced rapidly, which indicated that they were adapted to living and reproducing under the same conditions. But when they were cultured together, one strain out-competed and eventually eliminated the other. From this work Gause developed a fundamental concept in community ecology: the competitive exclusion principle, which states that if two competitors try to occupy the same realized niche, one species will eliminate the other. Many key questions about how species function in ecosystems can be answered by looking at their niches.

Species with narrow niches tend to be specialists, relying on comparatively few food sources. As a result, they are highly sensitive to changes in key environmental conditions, such as water temperature in aquatic ecosystems. For example, pandas, which only eat bamboo, have a highly specialized diet. Many endangered species are threatened because they live or forage in particular habitats that have been lost or converted to other uses.

Figure: Ecological niche

One well-known case, the northern spotted owl lives in cavities of trees in old-growth forests (forests with trees that are more than 200 years old and have not been cut, pruned, or managed), but these forests have been heavily logged, reducing the owl's habitat In contrast, species with broad niches are generalists that can adapt to wider ranges of environmental conditions within their own lifetimes (i.e., not through evolution over generations, but rather through changes in their behavior or physiologic functioning) and survive on diverse types of prey.

Coyotes once were found only on the Great Plains and in the western United States, but have spread through the eastern states in part because of their flexible lifestyle. They can kill and eat large, medium, or small prey, from deer to house cats, as well as other foods such as invertebrates and fruit, and can live in a range of habitats, from forests to open landscapes, farmland, and suburban neighbourhoods. Overlap between the niches of two species (more precisely, overlap between their resource use curves) causes the species to compete if resources are limited. One might expect to see species constantly dying off as a result, but in many cases competing species can coexist without either being eliminated. This happens through niche partitioning (also referred to as resource partitioning), in which two species divide a limiting resource such as light, food supply, or habitat.

Natural Ecosystem Change

Just as relationships between individual species are dynamic, so too is the overall makeup of ecosystems. The process by which one natural community changes into another over a time scale of years to centuries is called succession. Common succession patterns include plant colonization of sand dunes and the regrowth of forests on abandoned farmland.

While the general process is widely recognized, ecologists have offered differing views of what drives succession and how to define its end point. By analyzing the natural succession process, scientists seek to measure how stable ecosystems are at different stages in their trajectory of development, and how they respond to disturbances in their physical environment or changes in the frequency at which they are disturbed.

In the early 20th century, plant biologist Frederic Clements described two types of succession: primary (referring to colonization of a newly exposed landform, such as sand dunes or lava flows after a volcanic eruption) and secondary (describing the return of an area to its natural vegetation following a disturbance such as fire, treefall, or forest harvesting).

British ecologist Arthur Tansley distinguished between autogenic succession—change driven by the inhabitants of an ecosystem, such as forests regrowing on abandoned agricultural fields—and allogenic succession, or change driven by new external geophysical conditions such as rising average temperatures resulting from global climate change. As discussed above, ecologists often group species depending on whether they are better adapted for survival at low or high population densities (r-selected versus K-selected).

Succession represents a natural transition from r- to K-selected species. Ecosystems that have recently experienced traumatic extinction events such as floods or fires are favorable environments for r-selected species because these organisms, which are generalists and grow rapidly, can increase their populations in the absence of competition immediately after the event. Over time, however, they will be outcompeted by K-selected species, which often derive a competitive advantage from the habitat modification that takes place during early stages of primary succession.

For example, when an abandoned agricultural field transitions back to forest, sun-tolerant weeds and herbs appear first, followed by dense shrubs like hawthorn and blackberry. After about a decade, birches and other small fast-growing trees move in, sprouting wherever the wind blows their lightweight seeds. In 30 to 40 years, slower-spreading trees like ash, red maple, and oak take root, followed by shade-tolerant trees such as beech and hemlock. A common observation is that as ecosystems mature through successional stages, they tend to become more diverse and complex. The number of organisms and species increases and niches become narrower as competition for resources increases. Primary production rates and nutrient cycling may slow as energy moves through a longer sequence of trophic levels.

Many natural disturbances have interrupted the process of ecosystem succession throughout Earth's history, including natural climate fluctuations, the expansion and retreat of glaciers, and local factors such as fires and storms. An understanding of succession is central for conserving and restoring ecosystems because it identifies conditions that managers must create to bring an ecosystem back into its natural state. The Tallgrass Prairie National Preserve in Kansas, created in 1996 to protect 11,000 acres of prairie habitat, is an example of a conservation project that seeks to approximate natural ecosystem succession. A herd of grazing buffalo tramples on tree seedlings and digs up the ground, creating bare patches where new plants can grow, just as millions of buffalo maintained the grassland prairies that covered North America before European settlement.

Ecology

- **Definition and Scope of Ecology**

Human being has been interested in ecology in a practical sort of way since early in his history. In primitive society every individual, to survive, need to have definite knowledge of his environment, i.e., of the force of nature and of the plants and animals around him. Ecology is one of the popular areas of sciences in biology. It is a pluralistic science in the sense that it depends on a wide variety of methods and approaches rather than on a limited range of techniques and concepts. Even if, it is thought as part of biology, one important way in which ecology differs from most other branches of biology is that it can be properly appreciated or studied only through a multidisciplinary approach involving close cooperation from expertise in several disciplines.

- **Definition**

The word 'Ecology' was coined from the Greek word 'oikos' meaning 'house' or ' a place to live' to designate the study of organisms in their natural homes. Specially, it means the study of interactions of organisms with one another and with the physical and chemical environment. The term "logy" is to mean study. Another way of defining Ecology is to look at the levels of biological organizations. The molecules of life are organized in specific ways to form cells; cells are grouped in to tissues; and tissues are arranged to produce functional organs. The body organs are integrated to produce organ system, and the entire array of these systems constitutes an organism. Organisms exist not just as a single individual, but in-groups called population.

The various populations of organisms that interact with one another to form a community; interdependent communities of organisms interact with the physical environment to compose an ecosystem. Finally, all the ecosystems of the planet are combined to produce a level of organization known as the biosphere. Ecology is concerned with the levels of organization beyond that of individual organism; i.e. population, community, ecosystem, and biosphere.

- **Evolution of Definitions of Ecology**

Ecology = from the Greek root OIKOS, "at home", and OLOGY, "the study of"

Haeckle (1870): "By ecology we mean the body of knowledge concerning the economy of Nature - the investigation of the total relations of the animal to its inorganic and organic environment."

Burdon-Sanderson (1890s): Elevated Ecology to one of the three natural divisions of Biology: Physiology - Morphology –

Ecology Elton (1927): "Scientific natural history"

Andrewartha (1961): "The scientific study of the distribution and abundance of organisms"

Odum (1963): "The structure and function of Nature"

- **Scope**

Whether we are talking about humans or any other kind of organisms, certain principles govern the growth and stability of their populations over time. These principles influence the pattern of relationships of organisms with one another and their environment. These patterns, in all their varied forms, are the focuses of ecology. As a science, ecology seeks to treat the world of nature including its human component with a single set of concepts and principles.

Ecology deals with such questions as:

• Why natural communities are composed of certain organisms and not others;
• How the various organisms interact with each other and with the physical environment; and
• How we can control and maintain these natural communities

- **Human Activities Affecting Health and the Environment**

Human activities in an ecosystem have many drawbacks, unless we are approaching it in an environmental friendly way. The atmosphere, fertile soils, freshwater resources, the oceans and the ecosystems they support, play a key role in providing humans with shelter, food, safe water and the capacity to recycle most wastes. However, pressures exerted by humans, on the environment, in the form of pollution, resource depletion, land use changes and others affect environmental quality. Degradation of environmental quality can, in turn, lead to adverse human exposures and eventual health effects. The pressures excreted by the driving forces are in many instances increasing. They relate to household wastes, freshwater use, land use and agricultural development, industrialization and energy use. Household wastes Gaseous household wastes arise mainly from heating and cooking.

They contribute substantially to both outdoor and indoor air pollution. Liquid wastes are the by-products of domestic activities. In most areas of developing countries, feces are recycled for use in agriculture or deposited on land without prior destruction of pathogens. Not surprisingly, infectious disease such as diarrhoeal diseases, schistosomiasis and hepatitis are endemic, and some times epidemic, in such areas. Solid waste can also create environmental health problems.

It consists mainly of non-hazardous materials such as paper and plastic packaging material, glass, food scraps and other residues. However, it generally also contains small quantities of hazardous substances such as paints, medicines, solvents, cleaning materials and batteries, leading to potential chemical exposures. Production of household and municipal solid waste continues to increase worldwide, both in absolute and per capita terms.

Fresh Water

For a large percentage of the world's population, water supplies are neither safe nor adequate. Currently, over 1000 million people do not have access to an adequate supply of safe water for household consumption. Moreover, the world's freshwater resources are limited and unevenly distributed over the global land mass. Demand for water is nevertheless increasing in several sectors: for drinking water (domestic needs), food production (agriculture) and product manufacturing (industry). Global freshwater resources are threatened not only by overexploitation, however, but also by poor management and ecological degradation.

Untreated sewage is discharged into rivers and lakes; industrial wastes are dumped into water bodies; and runoff from agricultural fields treated with herbicides and pesticides is leading to water contamination. Industrial development, the exponential growth of human settlements and the ever-increasing use of synthetic organic substances are also having serious adverse impacts on freshwater bodies. Many surface and ground waters are now contaminated with nutrients, heavy metals and persistent organic pollutants. For instance, the River Awetu has been degraded by untreated liquid and solid waste discharge from Jimma Town, southwestern Ethiopia. The water is pungent and turns black just before the confluence point with the River Gilgel Gibe and no macroinvertebrates were found at this site (Worku et al., 2001).

Land use and Agricultural Development

Competition for land appears to be intensifying between sectors and production systems. Agriculture, in particular, can be expected to become an even more dominant form of land use. Population increases and the finite extent, to which further land can be converted to agricultural uses, mean that per capita arable land availability is becoming an issue. Agricultural production carries several risks. Thus extension and intensification of agricultural production systems, together with fluctuation in the supply of and demand for agricultural produce are causing shifts in the environmental determinants of the health status of local communities.

Erosion

Perhaps the worst erosion problem in the world, per ha of farmland, is in Ethiopia. Although Ethiopia has only 1/100 as much cropland in cultivation as the United States, it is thought to lose 2 billion metric tons of soil each year to erosion. Haitit is another country with severely degraded soil once covered with lush tropical forest; the land has been denuded for firewood and cropland. Erosion has been so bad that some experts now say the country has absolutely top soil, and poor peasant farmers have difficulty raising any crop at all. Economist Lester Brown of World Watch Institute warns that the country may never recover from this eco-disaster.

Industrialization

Industrialization is central to economic development and improved prospects for human well-being. But, if proper abatement technology is not used, industry becomes a major source of air, water and soil pollution, hazardous wastes and noise. Industrial workers are often at highest risk of health impacts. Furthermore, developed countries have exacerbated the environmental problems now being experienced by developing countries through transfer of hazardous wastes industries and technologies. Major industrial impacts also arise from small-scale industry. In developing countries, small-scale industry contributes substantially to economic development, but can create problems for environment and health if environmental safeguards are not used.

Energy

Energy plays a critical role in basic human survival. Energy has important implications for health. Energy is also crucial to transportation and industrial processes. However, production and use of energy, if not properly controlled may be accompanied by adverse health and environment impacts.

In developing countries, biomass accounts for about one-third of all energy use, and in some of least-developed countries, for as much as two-thirds. Open fires impair indoor air quality, add to the risk of accidents and jeopardize food hygiene. In general, the adverse effects on the environment of human activities are many and appear to be growing in intensity, and affecting larger and larger areas. Current and future potential pressures on the environment have major implication for health.

Environmental Threats to Human Health

Environmental threats to human health are numerous. These threats can be divided in to two:

a. Traditional hazard; i.e. associated with lack of development. Traditional hazards related to poverty and "insufficient" development are wide-ranging and include: lack of access to safe drinking-water; inadequate basic sanitation in the household and the community; indoor air pollution from cooking and heating using coal or biomass fuel and inadequate solid waste disposal.

b. Modern hazard, i.e. associated with unsustainable development. Modern hazards are related to development that lacks health- and environment safeguards, and to unsustainable consumption of natural resources. They include: water pollution from populated areas, industry and intensive agriculture; urban air pollution from motor cars, coal power stations and industry resulting in climate change, stratospheric ozone depletion and trans-boundary pollution.

Polluted air and water, excessive levels of noise, nuclear weapons fall-out, overcrowded slums, toxic waste dumps, inadequate or overly adequate diet, stress, food contaminants, medical X-rays, drugs, cigarettes, unsafe working conditions and other can be regarded as causative agents of environmental diseases. In short environmental diseases are those diseases that are introduced to the environment by man due to his careless behaviour.

Most environmentally induced diseases, unlike those caused by bacteria or other pathogens, are difficult to cure but theoretically simple to prevent. Remove the adverse environmental influence and the ailment will disappear. This is simply to say that by: - Preventing discharges of poisonous pollutants into water and food - Avoiding exposure to radiation - Keeping away from cigarette smoke - Avoiding synthetic food colouring or material One of the problems with environmental health concern is our limited knowledge on those toxic agents that are actually distributed over our earth, due to different activities by man in the ecosystem. For example, world-wide, there are about 10 million chemical compounds that have been synthesized thus far. But only one per cent is produced commercially and is regulated.

Beyond Fundamental Ecology

Applied Ecology: Using ecological principles to maintain conditions necessary for the continuation of present day life on earth.

Industrial Ecology: The design of the industrial infrastructure such that it consists of a series of interlocking "technological ecosystems" interfacing with global natural ecosystems. Industrial ecology takes the pattern and processes of natural ecosystems as a design for sustainability. It represents a shift in paradigm from conquering nature to becoming nature.

Ecological Engineering:

Unlike industrial ecology, the focus of Ecological Engineering is on the manipulation of natural ecosystems by humans for our purposes, using small amounts of supplemental energy to control systems in which the main energy drives are still coming from non-human sources. It is the design of new ecosystems for human purposes, using the self-organizing principles of natural ecosystems. [Note: The popular definition of ecological engineering is "the design of human society with its natural environment for the benefit of both.". What is the logical flaw in this definition?]

Ecological Economics:

Integrating ecology and economics in such a way that economic and environmental policies are reinforcing rather than mutually destructive.

Urban ecology:

For ecologists, urban ecology is the study of ecology in urban areas, specifically the relationships, interactions, types and numbers of species found in urban habitats. Also, the design of sustainable cities, urban design programs that incorporate political, infrastructure and economic considerations.

Conservartion Biology:

The application of diverse fields and disciplines to the conservation of biological diversity.

Restoration Biology:

Appllication of ecosystem ecology to the restoration of deteriorated landscapes in an attempt to bring it back to its original state as much as possible. Example, prarie grass.

Landscape Ecology:

"Landscape ecology is concerned with spatial patterns in the landscape and how they develop, with an emphasis on the role of disturbance, including human impacts" (Smith and Smith). It is a relatively new branch of ecology, that employs Global Information Systems. The goal is to predict the responses of different organisms to changes in landscape, to ultimately facilitate ecosystem management.

Levels of Studying Ecology

Biosphere: The earth's ecosystem interacting with the physical environment as a whole to maintain a steady state system intermediate in the flow of energy between the high energy input of the sun and the thermal sink of space (merges with atmosphere, lithosphere, hydrosphere…).

↓

Biome: Large scale areas of similar vegetation and climatic characteristics.

↓

Ecosystem: Set of organisms and abiotic components connected by the exchange of matter and energy (forest, lake, coastal ocean). Or, "the smallest units that can sustain life in isolation from all but atmospheric surroundings."

↓

Community: Interacting populations which significantly affect each other's distributions and abundance (intertidal, hot spring, wetland).

↓

Population: Group of interacting and interbreeding organisms ↓ Cell/Organism → Organelle → Molecule → Atom

Ecological Principles

Ecological principles build on ecological concepts (which are understood to be true) to draw key conclusions that can then guide human applications (section 3) aimed at conserving biodiversity.

<u>Principle 1</u>

Protection of species and species' subdivisions will conserve genetic diversity.

At the population level, the important processes are ultimately genetic and evolutionary because these maintain the potential for continued existence of species and their adaptation to changing conditions. In most instances managing for genetic diversity directly is impractical and difficult to implement.

The most credible surrogate for sustaining genetic variability is maintaining not only species but also the spatial structure of genetic variation within species (such as sub-species and populations). Maintenance of populations distributed across a species' natural range will assist in conserving genetic variability. This ensures the continuation of locally adapted genetic variants. Retaining a variety of individuals and species permits the adaptability needed to sustain ecosystem productivity in changing environments and can also beget further diversity (future adaptability).

This will be particularly important given climate change; for example, the genetic potential of populations at the northern edge of their range in B.C. may be particularly important to help facilitate species adaptation to changes. Species that are collapsing towards the edge (versus centre) of their range and disjunct populations (where a local population is disconnected from the continuous range of the species) are also particularly important to consider, given climate change, in order to conserve genetic diversity and enable adaptation.

Principle 2

Maintaining habitat is fundamental to conserving species.

A species habitat is the ecosystem conditions that support its life requirements. Our understanding of habitat is based on our knowledge of a species' ecology and how that determines where a species is known to occur or likely to occur. Habitat can be considered at a range of spatial and temporal scales that include specific microsites (e.g., occupied by certain invertebrates, bryophytes, some lichens), large heterogeneous habitats, or occupancy of habitat during certain time periods (e.g., breeding sites, winter range areas).

Therefore conserving habitat requires a multi-scale approach from regions to landscapes to ecosystems to critical habitat elements, features and structures.

Principle 3

Large areas usually contain more species than smaller areas with similar habitat.

The theory of island biogeography illustrates a basic principle that large areas usually contain more species than smaller areas with similar habitat because they can support larger and more viable populations. The theory holds that the number of species on an island is determined by two factors: the distance from the mainland and island size. These would affect the rate of extinction on the islands and the level of immigration. Other factors being similar (including distance to the mainland), on smaller islands the chance of extinction is greater than on larger ones. This is one reason why larger islands can hold more species than smaller ones. In the context of applying the theory more broadly, the "island" can be any area of habitat surrounded by areas unsuitable for the species on the island. Therefore a system of areas conserved for biodiversity that includes large areas can effectively support more viable populations.

Principle 4

All things are connected but the nature and strength of those connections vary.

Species play many different roles in communities and ecosystems and are connected by those roles to other species in different ways and with varying degrees of strength. It is important to understand key interactions. Some species (e.g., keystone species) have a more profound effect on ecosystems than others. Particular species and networks of interacting species have key, broad-scale ecosystem-level effects while others do not. The ways in which species interact vary in addition to the strengths of those interactions. Species can be predator and/or prey, mutualist or synergist.

Mutualist species provide a mutually beneficial association for each other such as fungi that colonize plant roots and aid in the uptake of soil mineral nutrients. Synergistic species create an effect greater than that predicted by the sum of effects each is able to create independently. The key issue is that it is important to determine which among the many interactions are the strong ones because those are the ones toward which attention needs to be directed.

Principle 5

Disturbances shape the characteristics of populations, communities, and ecosystems.

The type, intensity, frequency and duration of disturbances shape the characteristics of populations, communities and ecosystems including their size, shape and spatial relationships. Natural disturbances have played a key role in forming and maintaining natural ecosystems by influencing their structure including the size, shape and distribution of patches.

The more regions, landscapes, ecosystems and local habitat elements resemble those that were established from natural disturbances, the greater the probability that native species and ecological processes will be maintained. This approach can be strengthened by developing an improved understanding of how ecosystems respond to both natural and human disturbances, thus creating opportunities to build resilience in the system. For example, high frequency, low intensity fires have shaped ponderosa pine ecosystems while low frequency, high intensity fires have shaped lodgepole pine ecosystems. Maintaining these ecosystems means restoring fire and/or designing management practices such as harvesting to reduce the differences between a managed landscape and a landscape pattern created by natural disturbance. Since ecosystems can change dramatically at the site level due to natural disturbances, considering their composition and structure of habitats at the landscape-level may be more useful. For terrestrial ecosystems, this means taking into account:

- species composition;
- the amount and patch size distribution;
- the variety and proportion of seral stages of terrestrial habitat from young to old; and
- the diversity of within community structure (e.g., a variety of amounts of snags and coarse woody debris within forest stands).

It is important to recognize that for some less mobile species, distribution of habitat is potentially as influential as amount of habitat (i.e., patch size; connectivity).

Principle 6

Climate influences terrestrial, freshwater and marine ecosystems.

Climate is usually defined as all of the states of the atmosphere seen at a place over many years. Climate has a dominant effect on biodiversity as it influences meteorological variables like temperature, precipitation and wind with consequences for many ecological and physical processes, such as photosynthesis and fire behaviour. For example, major temperature fluctuations in surface waters in the

Pacific Ocean due to El Nino climatic events can influence weather and significantly warm temperatures throughout much of B.C. This in turn can increase some wildlife populations or impact the migration timing of some migratory bird populations. Another example of the effect of climate is the loss of large populations of native B.C. oysters due to cold temperatures in the 1900s; similarly, cold periods can kill fish in lakes. Because of the key role of climate, rapid climate change profoundly changes ecosystems.42 For example, climate change enables population outbreaks in some species and likely contributed to the mountain pine beetle epidemic in B.C. due to successive warm winters. Alterations to stream flow and timing of freshet resulting from climate change affects fish and waterfowl. A critical question therefore is: How should anticipated climate changes influence current conservation decisions so that ecosystems remain resilient in the future?

Unit-2

1.2 Nature, Scope and Content of Environmental Studies in Geography

Environment literally means Surrounding in which we are living. Environment includes all those things on which we are directly or indirectly dependent for our survival, whether it is living component like animals, plants or non-living component like soil, air water. Environmental studies are the scientific study of the environmental system and the status of its inherent or induced changes on organisms. It includes not only the study of physical and biological characters of the environment but also the social and cultural factors and the impact of man on environment.

Definitions of Environment

Some important definitions of environment are as under:

1. **Boring**: ‘A person’s environment consists of the sum total of the stimulation which he receives from his conception until his death.’ It can be concluded from the above definition that Environment comprises various types of forces such as physical, intellectual, economic, political, cultural, social, moral and emotional. Environment is the sum total of all the external forces, influences and conditions, which affect the life, nature, behaviour and the growth, development and maturation of living organisms.

2. **Douglas and Holland**: ‘The term environment is used to describe, in the aggregate, all the external forces, influences and conditions, which affect the life, nature, behaviour and the growth, development and maturity of living organisms.’

Elements of Environment

Environment is constituted by the interacting systems of physical, biological and cultural elements inter-related in various ways, individually as well as collectively. These elements may be explained as under:

(1) Physical elements

Physical elements are as space, landforms, water bodies, climate soils, rocks and minerals. They determine the variable character of the human habitat, its opportunities as well as limitations.

(2) Biological elements

Biological elements such as plants, animals, microorganisms and men constitute the biosphere.

(3) Cultural elements

Cultural elements such as economic, social and political elements are essentially manmade features, which make cultural milieu.

VARIOUS TYPES OF ENVIRONMENTS

According to Kurt Lewin, environment is of three types which influence the personality of an individual as under:

(a) Physical Environment,
(b) Social and Cultural Environment, and
(c) Psychological Environment.

These may be explained as under:

1. **Physical Environment**

Physical environment, refers to geographical climate and weather or physical conditions wherein and individual lives. The human races are greatly influenced by the climate. Some examples are as under: (a) In the cold countries i.e. European countries the people are of white colour. Likewise, in Asian and African countries, that is, in hot countries people are of dark complexion. (b) The physique of an individual depends on climate conditions as the individual tries to adjust in his physical environment. (d) The human working efficiency also depends on the climatic conditions.

2. **Social Environment**

Social Environment includes an individual's social, economic and political condition wherein he lives. The moral, cultural and emotional forces influence the life and nature of individual behaviour. Society may be classified into two categories as under: (i) An open society is very conductive for the individual developement. (ii) A closed society is not very conductive for the development.

3. **Psychological Environment**

Although physical and social environment are common to the individual in a specific situation. Yet every individual has his own psychological environment, in which he lives. Kurt Lewin has used the term 'life space' for explaining psychological environment. The Psychological environment enables us to understand the personality of an individual. Both the person and his goal form psychological environment. If a person is unable to overcome the barriers, he can either get frustrated or completed to change his goal for a new psychological environment. But adopting this mechanism, the individual is helped in his adjustment to the environment.

Scope and Importance of Environmental Studies:

The disciplines included in environmental education are environmental sciences, environmental engineering and environmental management.

(a) Environmental Science:

It deals with the scientific study of environmental system (air, water, soil and land), the inherent or induced changes on organisms and the environmental damages incurred as a result of human interaction with the environment.

(b) Environmental Engineering:

It deals with the study of technical processes involved in the protection of environment from the potentially deleterious effects of human activity and improving the environmental quality for the health and well beings of humans.

(c) Environmental Management:

It promotes due regard for physical, social and economic environment of the enterprise or projects. It encourages planned investment at the start of the production chain rather than forced investment in cleaning up at the end.

It generally covers the areas as environment and enterprise objectives, scope, and structure of the environment, interaction of nature, society and the enterprise, environment impact assessment, economics of pollution, prevention, environmental management standards etc.

The scope of environmental geography can be summed as follows:

- Environment geography is multi-disciplinary in nature. It is related to other disciplines like- life science, physical science, ecology, economics, biology, chemistry, public administration etc.
- The core is formed by the spatio-temporal relationship of man-environment explaining the role of physical factors in influencing the human activities as well as the human influence on the environment with the scientific and technological development.
- It is concerned with the spatial attributes of all the phenomena related to the environment.
- Studies the various biomes and human influences.
- Deals with the pattern of biodiversity at the global, national and local level.
- Studies the spatial pattern of physical and anthropogenic degradation of environment.
- Studies cause- effect,severity, management and mitigation of various environment issues like Climate change, global warming, ozone depletion, habitat loss, bio-diversity loss, pollution etc.
- Includes the notion of sustainable development, environment education, planning, conservation and management.

The importance of environmental studies is as follows:

1. To clarify modern environmental concept like how to conserve biodiversity.

2. To know the more sustainable way of living.

3. To use natural resources more efficiently.

4. To know the behaviour of organism under natural conditions.

5. To know the interrelationship between organisms in populations and communities.

6. To aware and educate people regarding environmental issues and problems at local, national and international levels.

Nature or environment sustains life. As a conscious and rational being, man needs to know the importance of environment and help keep the environment as healthy and productive as it can be. It is the environment that has made this beautiful world possible for him. Hence, there is an ever demanding need for environmental studies.

The natural environment that mankind had before the onset of industrialization, urbanization, and exponential growth in population was expectedly healthy and resilient. Nature was able to replenish the loss of its resources, which was very limited. After the onset of modern civilization, the overall health and efficiency of natural environment started deteriorating gradually and went on to such an extent that nature has virtually lost its natural ability to replenish the loss of resources caused by man. Environmentalists, geographers, and biologists the world over are constantly endeavouring for a sustainable solution to restore a sustainable environment.

There is a need for focus on environmental management, laws governing environment protection, pollution and recycling of non-bio-degradable material, etc. There is also a need for careful and cautious use of natural resources in the present time to establish sustainability in every aspect of nature. There is a need to clarify modern environmental concepts such as how to conserve biodiversity and maintain an ecological balance.

Environmental studies help us understand the importance of our environment and teaches us to use natural resources more efficiently and embrace a sustainable way of living. It enables us to know the behaviour of organisms under natural conditions and the interrelationship between organisms in population and communities.

According to UNESCO (1971), the objectives of environmental studies are:

(a) Creating the awareness about environmental problems among people.
(b) Imparting basic knowledge about the environment and its allied problems.
(c) Developing an attitude of concern for the environment.
(d) Motivating public to participate in environment protection and environment improvement.
(e) Acquiring skills to help the concerned individuals in identifying and solving environmental problems.
(f) Striving to attain harmony with Nature.

According to UNESCO, the guiding principles of environmental education should be as follows:

(a) Environmental education should be compulsory, right from the primary up to the post graduate stage.

(b) Environmental education should have an interdisciplinary approach by including physical, chemical, biological as well as socio-cultural aspects of the environment. It should build a bridge between biology and technology.

(c) Environmental education should take into account the historical perspective, the current and the potential historical issues.

(d) Environmental education should emphasise the importance of sustainable development i.e., economic development without degrading the environment.

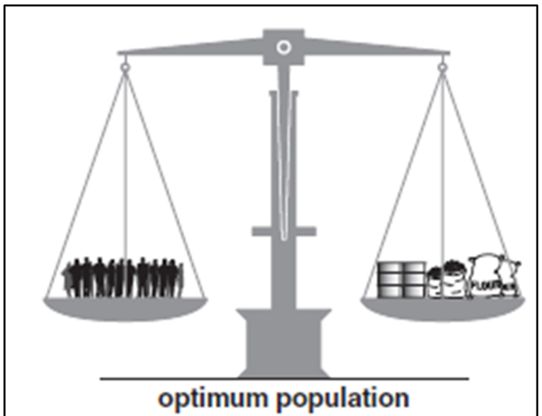
(e) Environmental education should emphasise the necessity of seeking international cooperation in environmental planning.

(f) Environmental education should lay more stress on practical activities and first hand experiences.

Unit-3

1.3 Concept of Optimum population

Optimum population has been defined as that size of population enabling per capita output of the maximum orders accompanied by the highest possible standards of living under a given set of economic and technological conditions. Therefore, optimum population lies between two extremes, i.e., overpopulation and under-population, although the size of optimum population is not sacrosanct.



https://sites.google.com/site/

It is a theoretically perfect situation difficult to estimate or define. The Penguin Dictionary of Geography characterises optimum population as a situation when the number of individuals can be accommodated in an area to the maximum advantage of each individual.

Thus optimum population yields highest quality of life, which means each person has access to adequate food, water, energy and air of highest quality, adequate medical care, recreational facilities and cultural outlets. In other words, optimum population permits the highest per capita output; therefore the marginal productivity exceeds the average productivity whereby the rates of growth of total production are the highest.

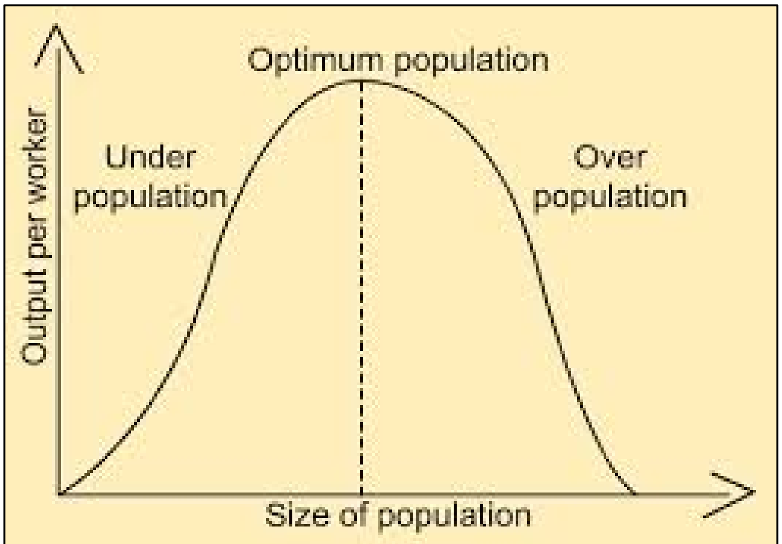


Figure: Concept of Over-population, Under-population and Optimum Population

Source: accessed from h://www.economicsdiscussion.net/

1.5 Gaia-hypothesis

- **Introduction**

The Gaia hypothesis is an ecological hypothesis proposing that the biosphere and the physical components of the Earth (atmosphere, cryosphere, hydrosphere and lithosphere) are closely integrated to form a complex interacting system that maintains the climatic and biogeochemical conditions on Earth in a preferred homeostasis. Originally proposed by James Lovelock as the earth feedback hypothesis, it was named the Gaia Hypothesis after the Greek supreme goddess of Earth. The hypothesis is frequently described as viewing the Earth as a single organism. Lovelock and other supporters of the idea now call it Gaia theory, regarding it as a scientific theory and not mere hypothesis, since they believe it has passed predictive tests.

- **History**

The Gaia hypothesis was first scientifically formulated in the 1960s by the independent research scientist James Lovelock, as a consequence of his work for NASA on methods of detecting life on Mars. He initially published the Gaia Hypothesis in journal articles in the early 1970s followed by a popularizing 1979 book Gaia: A new look at life on Earth . The theory was initially, according to Lovelock, a way to explain the fact that combinations of chemicals including oxygen and methane persist in stable concentrations in the atmosphere of the Earth. Lovelock suggested using such combinations detected in other planets' atmospheres would be a relatively reliable and cheap way to detect life, which many biologists opposed at the time and since.

Later other relationships such as the fact that sea creatures produce sulfur and iodine in approximately the quantities required by land creatures emerged and helped bolster the theory. Rather than invent many different theories to describe each such equilibrium, Lovelock dealt with them holistically, naming this self-regulating living system after the Greek goddess Gaia, using a suggestion from the novelist William Golding, who was living in the same Gaia Hypothesis village as Lovelock at the time (Bowerchalke, Wiltshire, UK).

The Gaia Hypothesis has since been supported by a number of scientific experiments and provided a number of useful predictions, and hence is properly referred to as the Gaia Theory. Since 1971, the noted microbiologist Dr. Lynn Margulis has been Lovelock's most important collaborator in developing Gaian concepts.

Until 1975 the hypothesis was almost totally ignored. An article in the New Scientist of February 15, 1975, and a popular book length version of the theory, published in 1979 as The Quest for Gaia , began to attract scientific and critical attention to the hypothesis. The theory was then attacked by many mainstream biologists. Championed by certain environmentalists and climate scientists, it was vociferously rejected by many others, both within scientific circles and outside them.

- **Lovelock's Initial Hypothesis**

James Lovelock defined Gaia as: a complex entity involving the Earth's biosphere, atmosphere, oceans, and soil; the totality constituting a feedback or cybernetic system which seeks an optimal physical and chemical environment for life on this planet. His initial hypothesis was that the biomass modifies the conditions on the planet to make conditions on the planet more hospitable – the Gaia Hypothesis properly defined this "hospitality" as a full homeostasis. Lovelock's initial hypothesis, accused of being teleological by his critics, was that the atmosphere is kept in homeostasis by and for the biosphere. Lovelock suggested that life on Earth provides a cybernetic, homeostatic feedback system operated automatically and unconsciously by the biota, leading to broad stabilization of global temperature and chemical composition.

With his initial hypothesis, Lovelock claimed the existence of a global control system of surface temperature, atmosphere composition and ocean salinity. His arguments were:

Gaia Hypothesis - The global surface temperature of the Earth has remained constant, despite an increase in the energy provided by the Sun.

- Atmospheric composition remains constant, even though it should be unstable.

- Ocean salinity is constant. Since life started on Earth, the energy provided by the Sun has increased by 25% to 30%; however the surface temperature of the planet has remained remarkably constant when measured on a global scale.

Furthermore, he argued, the atmospheric composition of the Earth is constant. The Earth's atmosphere currently consists of 79% nitrogen, 20.7% oxygen and 0.03% carbon dioxide. Oxygen is the second most reactive element after fluorine, and should combine with gases and minerals of the Earth's atmosphere and crust. Traces of methane (at an amount of 100,000 tonnes produced per annum) should not exist, as methane is combustible in an oxygen atmosphere. This composition should be unstable, and its stability can only have been maintained with removal or production by living organisms. Ocean salinity has been constant at about 3.4% for a very long time. Salinity stability is important as most cells require a rather constant salinity and do not generally tolerate values above 5%. Ocean salinity constancy was a long-standing mystery, because river salts should have raised the ocean salinity much higher than observed. Recently it was suggested that salinity may also be strongly influenced by seawater circulation through hot basaltic rocks, and emerging as hot water vents on ocean spreading ridges.

However, the composition of sea water is far from equilibrium, and it is difficult to explain this fact without the influence of organic processes. The only significant natural source of atmospheric carbon dioxide (CO_2) is volcanic activity, while the only significant removal is through the precipitation of carbonate rocks. In water, CO_2 is dissolved as a "carbonic acid," which may be combined with dissolved calcium to form solid calcium carbonate (limestone). Both precipitation and solution are influenced by the bacteria and plant roots in soils, where they improve gaseous circulation, or in coral reefs, where calcium carbonate is deposited as a solid on the sea floor.

Calcium carbonate can also be washed from continents to the sea where it is used by living organisms to manufacture carbonaceous tests and shells. Once dead, the living organisms' shells fall to the bottom of the oceans where they generate deposits of chalk and limestone. Part of the organisms with carbonaceous shells are the coccolithophores (algae), which also have a role in the formation of clouds. When they die, they release dimethyl sulfide gas (DMS), (CH_3), which is converted by atmospheric processes to sulfate particles on which water vapour condenses to make clouds. Lovelock sees this as one of the complex processes that maintain conditions suitable for life.

The volcanoes produce CO_2 in the atmosphere, CO_2 participates in rock weathering as carbonic acid, itself accelerated by temperature and soil life, the dissolved CO_2 is then used by the algae and released on the ocean floor. CO_2 excess can be compensated by an increase of coccolithophoride life, increasing the amount of CO_2 locked in the ocean floor. Coccolithophorides increase the cloud cover, hence control the surface temperature, help cool the whole planet and favour precipitations which are necessary for terrestrial plants. For Lovelock and other Gaia scientists like Stephan Harding, coccolithophorides are one stage in a regulatory feedback loop. Lately the atmospheric CO_2 concentration has increased and there is some evidence that concentrations of ocean algal blooms are also increasing.

- **Overview**

The Gaia theory posits that the Earth is a self-regulating complex system involving the biosphere, the atmosphere, the hydrospheres and the pedosphere, tightly coupled as an evolving system. The theory sustains that this system as a whole, called Gaia, seeks a physical and chemical environment optimal for contemporary life.

Gaia evolves through a cybernetic feedback system operated unconsciously by the biota, leading to broad stabilization of the conditions of habitability in a full homeostasis. Many processes in the Earth's surface essential for the conditions of life depend on the interaction of living forms, especially microorganisms, with inorganic elements. These processes establish a global control system that regulates Earth's surface temperature, atmosphere composition and ocean salinity, powered by the global thermodynamic desequilibrium state of the Earth system.

The existence of a planetary homeostasis influenced by living forms had been observed previously in the field of biogeochemistry, and it is being investigated also in other fields like Earth system science. The originality of the Gaia theory relies on the assessment that such homeostatic balance is actively pursued with the goal of keeping the optimal conditions for life, even when terrestrial or external events menace them.

- **Daisyworld Simulations**

Plots from a standard black & white DaisyWorld simulation. James Lovelock and Andrew Watson developed the mathematical model Daisyworld, that shows how temperature regulation can arise from organisms interacting with their environment. The purpose of the model is to demonstrate that feedback mechanisms can evolve from the actions or activities of self-interested organisms, rather than through classic group selection mechanisms.

Daisyworld examines the energy budget of a planet populated by two different types of plants, black daisies and white daisies. The colour of the daisies influences the albedo of the planet such that black daisies absorb light and warm the planet, while white daisies reflect light and cool the planet. Competition between the daisies (based on temperature-effects on growth rates) leads to a balance of populations that tends to favour a planetary temperature close to the optimum for daisy growth.

- **Regulation of the Salinity in the oceans**

Ocean salinity has been constant at about 3.4% for a very long time. Salinity stability in oceanic environments is important as most cells require a rather constant salinity and do not generally tolerate values above five per cent. Ocean salinity constancy was a long-standing mystery, because river salts should have raised the ocean salinity much higher than observed. Recently it was suggested that salinity may also be strongly influenced by seawater circulation through hot basaltic rocks, and emerging as hot water vents on mid-ocean ridges. However, the composition of seawater is far from equilibrium, and it is difficult to explain this fact without the influence of organic processes. One suggested explanation lies in the formation of salt plains throughout Earth's history. It is hypothesised that these are created by bacteria colonies that fix ions and heavy metals during life processes.

- **Regulation of oxygen in the Atmosphere**

Levels of gases in the atmosphere in 420,000 years of ice core data from Vostok, Antarctica research station. Current period is at the left. The atmospheric composition remains fairly constant providing the ideal conditions for contemporary life.

All the atmospheric gases other than noble gases present in the atmosphere are either made by organisms or processed by them. The Gaia theory states that the Earth's atmospheric composition is kept at a dynamically steady state by the presence of life. The stability of the atmosphere in Earth is not a consequence of chemical equilibrium like in planets without life.

Oxygen is the second most reactive element after fluorine, and should combine with gases and minerals of the Earth's atmosphere and crust. Traces of methane (at an amount of 100,000 tonnes produced per annum) should not exist, as methane is combustible in an oxygen atmosphere. Dry air in the atmosphere of Earth contains roughly (by volume) 78.09% nitrogen, 20.95% oxygen, 0.93% argon, 0.039% carbon dioxide, and small amounts of other gases including methane. While air content and atmospheric pressure varies at different layers, air suitable for the survival of terrestrial plants and terrestrial animals is currently known only to be found in Earth's troposphere and artificial atmospheres. Oxygen is a crucial element for the life of organisms, who require it at stable concentrations.

- **Regulation of the Global Surface Temperature**

Since life started on Earth, the energy provided by the Sun has increased by 25% to 30%; however, the surface temperature of the planet has remained within the levels of habitability, reaching quite regular low and high margins. Lovelock has also hypothesised that methanogens produced elevated levels of methane in the early atmosphere, giving a view similar to that found in petrochemical smog, similar in some respects to the atmosphere on Titan.

This, he suggests tended to screen out ultraviolet until the formation of the ozone screen, maintaining a degree of homeostasis. The Snowball Earth research, as a resultof "oxygen shocks" and reduced methane levels, that led during the Huronian, Sturtian and Marinoan/Varanger Ice Ages the world to very nearly become a solid "snowball" contradicts the Gaia hypothesis somewhat, although the ending of these Cryogenian periods through bio-geophysiological processes accords well with Lovelock's theory. Processing of the greenhouse gas CO_2 , explained below, plays a critical role in the maintenance of the Earth temperature within the limits of habitability.

The CLAW hypothesis, inspired by the Gaia theory, proposes a feedback loop that operates between ocean ecosystems and the Earth's climate. The hypothesis specifically proposes that particular phytoplankton that produce dimethyl sulphide are responsive to variations in climate forcing, and that these responses lead to a negative feedback loop that acts to stabilise the temperature of the Earth's atmosphere. Currently this Gaian homeostatic balance is being pushed by the increase of human population and the impact of their activities to the environment. The multiplication of greenhouse gases may cause a turn of Gaia's negative feedbacks into homeostatic positive feedback. According to Lovelock, this could bring an accelerated global warming and mass human mortality.

- **Biodiversity and Stability of Ecosystems**

The importance of the large number of species in an ecosystem, led to two sets of views about the role played by biodiversity in the stability of ecosystems in Gaia theory. In one school of thought labelled the "species redundancy" hypothesis, proposed by Australian ecologist Brian Walker, most species are seen as having little contribution overall in the stability, comparable to the passengers in an aeroplane who play little role in its successful flight. The hypothesis leads to the conclusion that only a few key species are necessary for a healthy ecosystem. The "rivet-popper" hypothes is put forth by Paul R. Ehrlich and his wife Anne H. Ehrlich, compares each species forming part of an ecosystem as a rivet on the aeroplane (represented by the ecosystem). The progressive loss of species mirrors the progressive loss of rivets from the plane, weakening it till it is no longer sustainable and crashes.

Later extensions of the Daisyworld simulation which included rabbits, foxes and other species, led to a surprising finding that the larger the number of species, the greater the improving effects on the entire planet (i.e., the temperature regulation was improved). It also showed that the system was robust and stable even when perturbed.

Daisyworld simulations where environmental changes were stable gradually became less diverse over time; in contrast gentle perturbations led to bursts of species richness. These findings lent support to the idea that biodiversity is valuable. This finding was later proved in a eleven-year old study of the factors species composition, dynamics and diversity in successional and native grasslands in Minnesota by David Tilman and John A. Downing wherein they discovered that "primary productivity in more diverse plant communities is more resistant to, and recovers more fully from, a major drought." They go on to add "Our results support the diversity stability hypothesis but not the alternative hypothesis that most species are functionally redundant."

- **Processing of CO_2**

Gaia scientists see the participation of living organisms in the Carbon cycle as one of the complex processes that maintain conditions suitable for life. The only significant natural source of atmospheric carbon dioxide (CO_2) is volcanic activity, while the only significant removal is through the precipitation of carbonate rocks.

Carbon precipitation, solution and fixation are influenced by the bacteria and plant roots in soils, where they improve gaseous circulation, or in coral reefs, where calcium carbonate is deposited as a solid on the sea floor. Calcium carbonate is used by living organisms to manufacture carbonaceous tests and shells. Once dead, the living organisms' shells fall to the bottom of the oceans where they generate deposits of chalk and limestone. One of these organisms is Emiliania huxleyi, an abundant coccolithophore algae which also has a role in the formation of clouds.

CO_2 excess is compensated by an increase of coccolithophoride life, increasing the amount of CO_2 locked in the ocean floor. Coccolithophorides increase the cloud cover, hence control the surface temperature, help cool the whole planet and favour precipitations necessary for terrestrial plants. Lately the atmospheric CO_2 concentration has increased and there is some evidence that concentrations of ocean algal blooms are also increasing. Lichen and other organisms accelerate the weathering of rocks in the surface, while the decomposition of rocks also happens faster in the soil, thanks to the activity of roots, fungi, bacteria and subterranean animals. The flow of carbon dioxide from the atmosphere to the soil is therefore regulated with the help of living beings. When CO2 levels rise in the atmosphere the temperature increases and plants grow. This growth brings higher consumption of CO2 by the plants, who process it into the soil, removing it from the atmosphere.

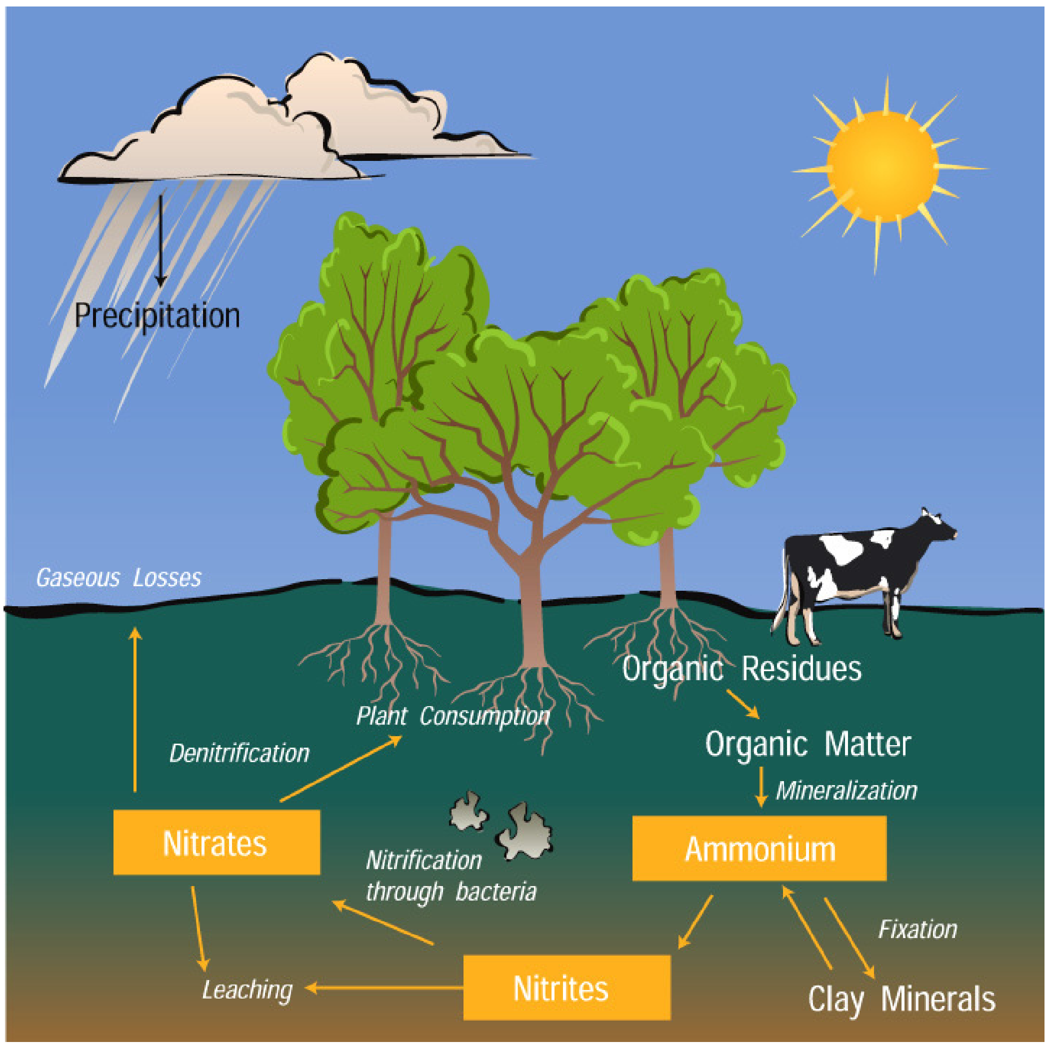


Figure: Model of Gaia Hypothesis

- **Criticisms**

After initially being largely ignored by most scientists, (from 1969 until 1977), thereafter for a period, the initial Gaia hypothesis was ridiculed by a number of scientists, such as Ford Doolittle, Richard Dawkins and Stephen Jay Gould. Lovelock has said that by naming his theory after a Greek goddess, championed by many non-scientists, the Gaia hypothesis was derided as some kind of neo-Pagan New Age religion. Many scientists in particular also criticised the approach taken in his popular book "Gaia, a New look at Life on Earth" for being teleological; a belief that all things have a predetermined purpose. Responding to this assertion in 1990, Lovelock stated "Nowhere in our writings do we express the idea that planetary self-regulation is purposeful, or involves foresight or planning by the biota." Stephen Jay Gould criticised Gaia as merely a metaphorical description of Earth processes. He wanted to know the actual mechanisms by which self-regulating homeostasis was regulated. David Abram argued that Gould was unaware that mechanism was itself only metaphorical.

Lovelock argues that no one mechanism is responsible, that the connections between the various known mechanisms may never be known, that this is accepted in other fields of biology and ecology as a matter of course, and that specific hostility is reserved for his own theory for other reasons. Aside from clarifying his language and understanding of what is meant by a life form, Lovelock himself ascribes most of the criticism to a lack of understanding of non-linear mathematics by his critics, and a linearizing form of greedy reductionism in which all events have to be immediately ascribed to specific causes before the fact. He notes also that his theory suggests experiments in many different fields, but few of them in biology, which most of his critics are trained in. "I'm a general practitioner in a world where there's nothing but specialists... science in the last two centuries has tended to be ever-dividing" and often rivalrous, especially for funding, which Lovelock describes as overly abundant and overly focused on institutions rather than original thought. He points out that Richard Feynman not only shared this opinion (coining the term cargo cult science) but also accepted a lack of general causeand effect explanation as an inevitable phase in a theory's development, and believed that some self-regulating phenomena may not be explainable at all mathematically.

- **The Earth Alive**

The concept of a living Earth has caused a lot of controversy, partly due to the different attributes and connotations given to this hypothetical life, partly because of the straightforward language used by Lovelock in his writings. For instance, evolutionary biologists such as the late palaeontologist Stephen Jay Gould and the ethologist Richard Dawkins attacked his statement in the first paragraph of his book (1979), that "the quest for Gaia is an attempt to find the largest living creature on Earth."

James Lovelock sustains that agreeing on a rational answer is not possible because science has not yet formulated a full definition of life. A basic criterion of the empirical definition of a life-form is its birth out of natural selection and its ability to replicate and pass on its genetic information to a succeeding generation. Dawkins stressed that, consequently, an argument against the idea that Gaia as a living organism is the fact that the planet is not the offspring of any parents and is unable to reproduce. Lovelock, however, defines life as a self-preserving, self-similar system of feedback loops like Humberto Maturana's autopoiesis; as a self-similar system, life could be a cell as well as an organ embedded into a larger organism as well as an individual in a larger inter-dependent social context. The biggest context of interacting inter-dependent living entities is the Earth.

The problematic empirical definition is getting "fuzzy on the edges": Why are highly specialized bacteria, such as E. coli, unable to thrive outside their habitat considered "life", while mitochondria, which have evolved independently from the rest of the cell, are not? William Irwin Thompson suggests that the Chilean biologist Humberto Maturana and James Lovelock, with the deductive definition of autopoiesis, have provided an explanation for the phenomenon of life. Reproduction becomes optional: bee swarms reproduce, while the biosphere has no need to. Lovelock himself states in the original Gaia book that even that is not true; given the possibilities, the biosphere may multiply in the future by colonizing other planets, as humankind may be the primer by which Gaia will reproduce. Humanity's exploration of space, its interest in colonizing and even terraforming other planets, lends some plausibility to the idea that Gaia might in effect be able to reproduce.

- **Natural Selection and Evolution**

Lovelock accepts a process of systemic Darwinian evolution for such biological feedback mechanisms: creatures that improve their environment for their survival do better than those that damage their environment. However, some scientists dispute the existence of such mechanisms. In 1981, W. Ford Doolittle, in the Co-Evolution Quarterly article "Is Nature Motherly" argued that nothing in the genome of individual organisms could provide the feedback mechanisms Gaia theory proposed, and therefore the Gaia hypothesis was an unscientific theory of a maternal type without any explanatory mechanism. In Richard Dawkins' 1982 book, The Extended Phenotype, he argued that organisms could not act in concert as this would require foresight and planning from them. Like Doolittle he rejected the possibility that feedback loops could stabilize the system. Lynn Margulis, a microbiologist who collaborated with Lovelock in supporting the Gaia hypothesis, argued in 1999, that "Darwin's grand vision was not wrong, only incomplete.

In accentuating the direct competition between individuals for resources as the primary selection mechanism, Darwin (and especially his followers) created the impression that the environment was simply a static arena." She wrote that the composition of the Earth's atmosphere, hydrosphere, and lithosphere are regulated around "set points" as in homeostasis, but those set points change with time. She also wrote that there is no special tendency of biospheres to preserve their current inhabitants, and certainly not to make them comfortable. According to her, the Earth is a kind of community of trust that can exist at many discrete levels of integration. All multicellular organisms do not live or die all at once: not all cells in the body die

- **Range of views**

Some have found James Kirchner's suggested spectrum, proposed at the First Gaia Chapman Conference, useful in suggesting that the original Gaia hypothesis could be split into a spectrum of hypotheses, ranging from the undeniable (Weak Gaia) to the radical (Strong Gaia).

Weak Gaia

At one end of this spectrum is the undeniable statement that the organisms on the Earth have altered its composition. A stronger position is that the Earth's biosphere effectively acts as if it is a self-organizing system, which works in such a way as to keep its systems in some kind of "meta-equilibrium" that is broadly conducive to life. The history of evolution, ecology and climate show that the exact characteristics of this equilibrium intermittently have undergone rapid changes, which are believed to have caused extinctions and felled civilizations (see climate change). Weak Gaian hypotheses suggest that Gaia is co-evolutive.

Co-evolution in this context has been thus defined: "Biota influence their abiotic environment, and that environment in turn influences the biota by Darwinian process." Lovelock (1995) gave evidence of this in his second book, showing the evolution from the world of the early thermo-acido-philic and methanogenic bacteria towards the oxygen enriched atmosphere today that supports more complex life. The weakest form of the theory has been called "influential Gaia".

It states that biota minimally influence certain aspects of the abiotic world, e.g. temperature and atmosphere. The weak versions are more acceptable from an orthodox science perspective, as they assume non-homeostasis. They state the evolution of life and its environment may affect each other. An example is how the activity of photosynthetic bacteria during Precambrian times have completely modified the Earth atmosphere to turn it aerobic, and as such supporting evolution of life (in particular eukaryotic life).

However, these theories do not claim the atmosphere modification has been done in coordination and through homeostasis. Also such critical theories have yet to explain how conditions on Earth have not been changed by the kinds of run-away positive feedbacks that have affected Mars and Venus. Biologists and earth scientists usually view the factors that stabilize the characteristics of a period as an undirected emergent property or entelechy of the system; as each individual species pursues its own self-interest, for example, their combined actions tend to have counterbalancing effects on environmental change.

Opponents of this view sometimes reference examples of lives' actions that have resulted in dramatic change rather than stable equilibrium, such as the conversion of the Earth's atmosphere from a reducing environment to an oxygen-rich one. However, proponents argue these atmospheric changes improved the environment's suitability for life. Some go a step further and hypothesize that all life-forms are part of one single living planetary being called Gaia. In this view, the atmosphere, the seas and the terrestrial crust would be results of interventions carried out by Gaia through the coevolving diversity of living organisms. While it is arguable that the Earth as a unit does not match the generally accepted biological criteria for life itself (Gaia has not yet reproduced, for instance; it still might spread to other planets through human space colonization and terraforming), many scientists would be comfortable characterizing the earth as a single "system".

Strong Gaia

A version called "Optimizing Gaia" asserts that biota manipulate their physical environment for the purpose of creating biologically favorable, or even optimal, conditions for themselves. "The Earth's atmosphere is more than merely anomalous; it appears to be a contrivance specifically constituted for a set of purposes". Further, "... it is unlikely that chance alone accounts for the fact that temperature, pH and the presence of compounds of nutrient elements have been, for immense periods, just those optimal for surface life. Rather, ... energy is expended by the biota to actively maintain these optima". Another strong hypothesis is the one called "Omega Gaia".

Teilhard de Chardin claimed that the Earth is evolving through stages of cosmogenesis, affecting the geosphere, biogenesis of the biosphere, and noogenesis of the noosphere, culminating in the Omega Point. Another form of the strong Gaia hypothesis is proposed by Guy Murchie who extends the quality of a holistic lifeform to galaxies. "After all, we are made of star dust. Life is inherent in nature." Murchie describes geologic phenomena such as sand dunes, glaciers, fires, etc. as living organisms, as well as the life of metals and crystals. "The question is not whether there is life outside our planet, but whether it is possible to have "nonlife". There are speculative versions of the Gaia hypothesis, including versions that hold that the Earth is conscious or part of some universe-wide evolution such as expressed in the Selfish Biocosm hypothesis strain of a larger speculative Gaia philosophy.

These extreme forms of the Gaia hypothesis, that the entire Earth is a single unified organism that is consciously manipulating the climate to make conditions more conducive to life, are metaphysical or mystical views for which no evidence exists, and that cannot be tested scientifically. The political branch of Gaia theory is the Gaia Movement, a collection of different organisations operating in different countries, but all sharing a concern for how humans might live more sustainably within the "living system". History Precedents "Earthrise" taken on December 24, 1968. The holistic idea of the Earth as an integrated whole, a living being, has a long tradition. The mythical Gaia was the primal Greek goddess personifying the Earth, the Greek version of "Mother Nature," or the Earth Mother. James Lovelock gave this name to his hypothesis after a suggestion from the novelist William Golding, who was living in the same village as Lovelock at the time (Bowerchalke, Wiltshire, UK). Golding's advice was based on Gea, an alternative spelling for the name of the Greek goddess, which is used as prefix in geology, geophysics and geochemistry. In the eighteen century, as Geology consolidated as a modern science, James Hutton maintained that geological and biological processes are interlinked.

The naturalist and explorer Alexander von Humboldt recognized the coevolution of living organisms, climate, and Earth crust. Already in the twentieth century, Vladimir Vernadsky developed theory of the Earth's development that is now one of the foundations of Ecology. The Ukrainian geochemist was one of the first scientists to recognize that the oxygen, nitrogen and carbon dioxide in the Earth's atmosphere result from biological processes. During the 1920s he published works arguing that living organisms could reshape the planets as surely as any physical force. Vernadsky was an important pioneer of the scientific bases for the environmental sciences.

His visionary pronouncements were not widely accepted in the West, and some decades after the Gaia hypothesis received the same type of initial resistance from the scientific community. Also in the turn to the 20th century Aldo Leopold, pioneer in the development of modern environmental ethics and in the movement for wilderness conservation, suggested a living Earth in his bio-centric or holistic ethics regarding land.

> It is at least not impossible to regard the earth's parts—soil, mountains, rivers, atmosphere etc,— as organs or parts of organs of a coordinated whole, each part with its definite function. And if we could see this whole, as a whole, through a great period of time, we might perceive not only organs with coordinated functions, but possibly also that process of consumption as replacement which in biology we call metabolism, or growth. In such case we would have all the visible attributes of a living thing, which we do not realize to be such because it is too big, and its life processes too slow. — Quoted by Stephan Harding in Animate Earth.

Another influence for the Gaia theory and the environmental movement in general came as a side effect of the Space Race between the Soviet Union and the United States of America. During the 1960s, the first humans in space could see how the Earth looked alike as a whole. The photograph Earthrise taken by astronaut William Anders in 1968 during the Apollo 8 mission became an early symbol for the global ecology movement.

Unit-4

1.6 Deep Ecology

Norwegian philosopher Arne Naess (b. 1912) coined the term "Deep Ecology" in 1972 to express the idea that nature has intrinsic value, namely, value apart from its usefulness to human beings, and that all life forms should be allowed to flourish and fulfil their evolutionary destinies. Naess invented the rubric to contrast such views with what he considered to be "shallow" environmentalism, namely, environmental concern rooted only in concern for humans. The term has since come to signify both its advocates' deeply felt spiritual connections to the Earth's living systems and ethical obligations to protect them, as well as the global environmental movement that bears its name. Moreover, some deep ecologists posit close connections between certain streams in world religions and deep ecology.

Naess and most deep ecologists, however, trace their perspective to personal experiences of connection to and wholeness in wild nature, experiences which are the ground of their intuitive, affective perception of the sacredness and interconnection of all life. Those who have experienced such a transformation of consciousness (experiencing what is sometimes called one's "ecological self" in these movements) view the self not as separate from and superior to all else, but rather as a small part of the entire cosmos. From such experience flows the conclusion that all life and even ecosystems themselves have inherent or intrinsic value – that is, value independently of whether they are useful to humans.

Although Naess coined the term, many deep ecologists credit the American ecologist Aldo Leopold with succinctly expressing such a deep ecological worldview in his now famous "Land Ethic" essay, which was published posthumously in A Sand County Almanac in 1948. Leopold argued that humans ought to act only in ways designed to protect the long-term flourishing of all ecosystems and each of their constituent parts. Many deep ecologists call their perspective alternatively "ecocentrism" or "biocentrism" (to convey, respectively, an ecosystem-centered or life-centered value system). As importantly, they believe humans have so degraded the biosphere that its life-sustaining systems are breaking down. They trace this tragic situation to anthropocentrism (human-centeredness), which values nature exclusively in terms of its usefulness to humans.

Anthropocentrism, in turn, is viewed as grounded in Western religion and philosophy, which many deep ecologists believe must be rejected (or a deep ecological transformation of consciousness within them must occur) if humans are to learn to live sustainably on the Earth. Thus, many deep ecologists believe that only by "resacralizing" our perceptions of the natural world can we put ecosystems above narrow human interests and learn to live harmoniously with the natural world, thereby averting ecological catastrophe. It is a common perception within the deep ecology movement that the religions of indigenous cultures, the world's remnant and newly revitalized or invented pagan religions, and religions originating in Asia (especially Daoism, Buddhism, and Hinduism) provide superior grounds for ecological ethics, and greater ecological wisdom, than do Occidental religions.

Theologians such as Matthew Fox and Thomas Berry, however, have shown that Western religions such as Christianity may be interpreted in ways largely compatible with the deep ecology movement. Although Naess coined the umbrella term, which is now a catchphrase for most non-anthropocentric environmental ethics, a number of Americans were also criticizing anthropocentrism and laying the foundation for the movement's ideas at about the same time as Naess was coining the term. One crucial event early in deep ecology's evolution was the 1974 "Rights of Non-Human Nature" conference held at a college in Claremont, California. Inspired by Christopher Stone's influential 1972 law article (and subsequent book) Should Trees Have Standing? – Toward Legal Rights for Natural Objects, the conference drew many of those who would become the intellectual architects of deep ecology.

These included George Sessions who, like Naess, drew on Spinoza's pantheism, later coauthoring Deep Ecology with Bill Devall; Gary Snyder, whose remarkable, Pulitzer prize-winning Turtle Island proclaimed the value of place-based spiritualities, indigenous cultures, and animistic perceptions, ideas that would become central within deep ecology subcultures; and the late Paul Shepard (d. 1996), who in The Tender Carnivore and the Sacred Game, and subsequent works such as Nature and Madness and the posthumously published Coming Back to the Pleistocene, argued that foraging societies were ecologically superior to and emotionally healthier than agricultural societies. Shepard and Snyder especially provided a cosmogony that explained humanity's fall from a pristine, natural paradise. Also extremely influential was Edward Abbey's Desert Solitaire, which viewed the desert as a sacred place uniquely able to evoke in people a proper, non-anthropocentric understanding of the value of nature. By the early 1970s the above figures put in place the intellectual foundations of deep ecology.

A corresponding movement soon followed and grew rapidly, greatly influencing grassroots environmentalism, especially in Europe, North America, and Australia. Shortly after forming in 1980, for example, leaders of the politically radical Earth First! movement (the exclamation point is part of its name) learned about Deep Ecology, and immediately embraced it as their own spiritual philosophy. Meanwhile, the green lifestyle-focused movement known as bioregionalism also began to embody a deep ecology worldview.

Given their natural affinities it was not long before bioregionalism became the prevailing social philosophy among deep ecologists. As a philosophy and as a movement, deep ecology spread in many ways. During the 1980s and early 1990s, for example, Bill Devall and George Sessions published their influential book, Deep Ecology: Living as if Nature Mattered; Warwick Fox in Toward a Transpersonal Ecology linked deep ecology with transpersonal psychology, thereby furthering the development of what is now called ecopsychology; David Rothenberg translated and edited Arne Naess's important work, Ecology, Community and Lifestyle; and Michael E. Zimmerman interpreted Martin Heidegger as a forerunner of deep ecology, thus helping to spark a trend of calling upon contemporary European thinkers for insight into environmental issues.

Many deep ecologists have complained, however, that the postmodern thinking imported from Europe has undermined the status of "nature," defined by deep ecologists as a whole that includes but exists independently of humankind. Radical environmentalist activists, including the American co-founder of Earth First!, Dave Foreman, and the Australian co-founder of the Rainforest Information Centre, John Seed, beginning in the early 1980s, conducted "road shows" to transform consciousness and promote environmental action. Such events usually involve speeches and music designed to evoke or reinforce peoples' felt connections to nature, and inspires action. Often, they also include photographic presentations contrasting intact and revered ecosystems with degraded and defiled lands.

Some activists have designed ritual processes to further deepen participants' spiritual connections to nature and political commitment to defend it. Joanna Macy and a number of others, including John Seed, for example, developed a ritual process known as the Council of All Beings, which endeavors to get activists to see the world from the perspective of nonhuman entities. Since the early 1980s, traveling widely around the world, Seed has laboured especially hard spreading deep ecology through this and other newly invented ritual processes. The movement has also been disseminated through the writings of its architects (often reaching college students in environmental studies courses); through journalists reporting on deep ecology-inspired environmental protests and direct action resistance; and through the work of novelists, poets, musicians, and other artists, who promote in their work deep ecological perceptions. Recent expressions in ecotourism can be seen, for example, in the "Deep Ecology Elephant Project," which includes tours in both Asia and Africa, and suggest that elephants and other wildlife have much to teach their human kin.

Deep Ecology has been criticized by people representing social ecology, socialist ecology, liberal democracy, and ecofeminism. Murray Bookchin, architect of the anarchistic green social philosophy known as Social Ecology, engaged in sometimes vituperative attacks ondeep ecology and its activist vanguard, Earth First!, for being intellectually incoherent, ignorant of socioeconomic factors in environmental problems, and given to mysticism and misanthropy. Bookchin harshly criticized Earth First! co-founder Dave Foreman for suggesting that starvation was a solution to human overpopulation and environmental deterioration. Later, however, Bookchin and Foreman engaged in a more constructive dialogue. Meanwhile, socialist ecologists maintain that deep ecology overemphasizes cultural factors (worldviews, religion, philosophy) in diagnosing the roots of, and solutions to, environmental problems, thereby minimizing the roles played by the social, political, and economic factors inherent in global capitalism. Liberal democrats such as the French scholar Luc Ferry (1995) maintain that deep ecology is incapable of providing guidance in moral decision making. Insofar as deep ecology fails adequately to recognize that human life has more value than other life forms, he argues, it promotes "ecofascism," namely the sacrifice of individual humans for the benefit of the ecological whole, what Leopold termed "the land." (Ecofascism in its most extreme form links the racial purity of a people to the well-being of the nation's land; calls for the removal or killing of nonnative peoples; and may also justify profound individual and collective sacrifice of its own people for the health of the natural environment.)

Many environmental philosophers have defended Leopold's land ethic, and by extension, deep ecology, against such charges, most notably one of the pioneers of contemporary environmental philosophy, J. Baird Callicott. Although some ecofeminists indicate sympathy with deep ecology's basic goal, namely, protecting natural phenomena from human destruction, others have sharply criticized deep ecology. Male, white, and middle-class deep ecologists, Ariel Salleh maintains, ignore how patriarchal beliefs, attitudes, practices, and institutions help to generate environmental problems. Val Plumwood and Jim Cheney criticize deep ecology's idea of expanding the self so as to include and thus to have a basis for protecting nonhuman phenomena. This "ecological self" allegedly constitutes a totalizing view that obliterates legitimate distinctions between self and other. Moreover, Plumwood argues, deep ecology unwisely follows the rationalist tradition in basing moral decisions on "impartial identification," a practice that does not allow for the highly particular attachments that often motivate environmentalists and indigenous people alike to care for local places.

Warwick Fox has replied that impartial and wider identification does not cancel out particular or personal attachments, but instead, puts them in the context of more encompassing concerns that are otherwise ignored, as when for example concern for one's family blinds one to concerns about concerns of the community. Fox adds that deep ecology criticizes the ideology – anthropocentrism – that has always been used to by social agents to legitimate oppression of groups regarded as sub- or nonhuman. While modern liberation movements have sought to include more and more people into the class of full humans, such movements have typically not criticized anthropocentrism as such.

Even a fully egalitarian society, in other words, could continue to use anthropocentrism to justify exploiting the nonhuman realm. In response to the claim that deep ecology is, or threatens to be, a totalizing worldview that excludes alternatives and that – ironically – threatens cultural diversity, Arne Naess responds that, to the contrary, deep ecology is constituted by multiple perspectives or "ecosophies" (ecological philosophies) and is compatible with a wide range of religious perspectives and philosophical orientations. Another critic, best-selling author Ken Wilber, argues that by portraying humankind as merely one strand in the web of life, deep ecology adheres to a one-dimensional, or "flatland" metaphysics (1995).

Paradoxically, by asserting that material nature constitutes the whole of which humans are but a part, deep ecologists agree in important respects with modern naturalism, according to which humankind is a clever animal capable of and justified in dominating other life forms in the struggle for survival and power. According to Wilber, a "deeper" ecology would discern that the cosmos is hierarchically ordered in terms of complexity, and that respect and compassion are due all phenomena because they are manifestations of the divine. In the last analysis, for Naess, it is personal experiences of a profound connection with nature and related perceptions of nature's inherent worth or sacredness, which give rise to deep ecological commitments. Naess believes such commitments may be derived from a wide variety of ultimate premises, religious and philosophical, so as to form a particular ecosophy. Ecosophies that identify themselves as part of the Deep Ecology Movement are consistent with the eight-point, Deep Ecology Platform, which Naess developed with George Sessions in 1984.

Although controversial and contested, both internally and among its proponents and its critics, deep ecology is an increasingly influential green spirituality and ethics that is universally recognized in environmentalist enclaves, and increasingly outside of such subcultures, as a radical movement challenging the conventional, usually anthropocentric ways humans deal with the natural world. Its influence in environmental philosophy has been profound, for even those articulating alternative environmental ethics are compelled to respond to its insistence that nature has intrinsic and even sacred value, and its challenge to anthropocentrism. Its greatest influence, however, may be through the diverse forms of environmental activism that it inspires, action that increasingly shapes world environmental politics.

Not only is deep ecology the prevailing spirituality of bioregionalism and radical environmentalism; it also undergirds the International Forum on Globalization and the Ruckus Society, two organizations playing key roles in the anti-globalization protests that erupted in 1999. Both of these groups are generously funded by the San Francisco-based Foundation for Deep Ecology, and other foundations, which share deep ecological perceptions. Such developments reflect a growing impulse toward institutionalization, which is designed to promote deep ecology and intensify environmental action. There are now Institutes for Deep Ecology in London, England and Occidental, California, a Sierra Nevada Deep Ecology Institute in Nevada City, California, and dozens of other organizations in the United States, Oceania, and Europe, which provide ritual-infused experiences in deep ecology and training for environmental activists. It is not, however, the movement's institutions, but instead the participants' love for the living Earth, along with their widespread apocalypticism (their conviction that the world as we know it is imperiled or doomed), that give the movement its urgent passion to promote earthen spirituality, sustainable living, and environmental activism.

The Eight-Point Platform

Growing out of the knowledge of nature's concrete contents is the recognition of the need for some kind of political action. To this end Naess and Sessions laid out an oft-cited eight-point program (that they conjured while camping in Death Valley in 1984) For example (Naess 1986, p. 24), in the diagram Buddhist, secular philosophical, and Christian first principles (the bust) converge in the eight-point platform (the waist), which then justifies an array of activisms (the skirt [see Figure). Buddhist metaphysics might channel through the waist of deep ecological principles calling for environmental action toreduce consumption; secular metaphysics might channel through the waist of Deep Ecology calling for action to reduce human population growth; or Christian metaphysics might channel through the waist of Deep Ecology to call for action to preserve biodiversity. Both the eight-point platform and the apron diagram imply that Deep Ecology is above all an ontology and incidentally an ethic.

THE BASIC PRINCIPLES OF DEEP ECOLOGY

1. **Inherent value**

The well-being and flourishing of human and nonhuman Life on Earth have value in themselves (synonyms: intrinsic value, inherent value). These values are independent of the usefulness of the nonhuman world for human purposes.

2. **Diversity**

Richness and diversity of life forms contribute to the realization of these values and are also values in themselves.

3. **Vital Needs**

Humans have no right to reduce this richness and diversity except to satisfy vital needs.

4. **Population**

The flourishing of human life and cultures is compatible with a substantial decrease of the human population. The flourishing of nonhuman life requires such a decrease.

5. **Human Interference**

The present human interference with the nonhuman world is excessive, and the situation is rapidly worsening.

6. **Policy Change**

Policies must therefore be changed. These policies affect basic economic, technological, and ideological structures. The resulting state of affairs will be deeply different from the present.

7. **Quality of Life**

The ideological change is mainly that of appreciating life quality (dwelling in situations of inherent value) rather than adhering to an increasingly higher standard of living. There will be a profound awareness of the difference between big and great.

8. **Obligation of Action**

Those who subscribe to the foregoing points have an obligation directly or indirectly to try to implement the necessary changes.

CRITIQUES OF DEEP ECOLOGY

The deep-ecological principles of biocentric egalitarianism and metaphysical holism have elicited robust critiques. Some of the most interesting debates have centered on the normative status of Deep Ecology. Naess maintains that Deep Ecology is essentially descriptive. For Naess unmitigated empiricism or ‘‘ecophenomenology’’ (Brown and Toadvine 2003) promotes a direct experience of the qualities of nature—its ‘‘concrete contents’’ (Naess 1985). Deep Ecology, he argues, is simply an enumeration of general principles that command the assent of persons open to the direct apprehension of nature.

Scholars have found the disclaimer that Deep Ecology is not a normative system—and ought not be judged as such—disingenuous. They have treated Deep Ecology as the legitimate object of the analysis of moral philosophy. Some regard Deep Ecology as strident axiological egalitarianism that is useless in adjudicating conflicting interests. If all organisms are of equal value, then there is no basis upon which to make prescriptions because the kind of value distinctions necessary for evaluating the moral situations of environmental ethics are deliberately disqualified. The principle of bio-centric egalitarianism, on this view, renders Deep Ecology impotent as an ethical theory.

Environmental ethics is predicated on the possibility of a nonegalitarian axiology. In the words of the American philosopher Bryan Norton, ‘‘The 120,000th elk cannot be treated equally with one of the last California condors—not, at least, on a reasonable environmental ethic’’ (1991, p. 224). Baird Callicott has surmised that environmental ethics must manifestly not ‘‘accord equal moral worth to each and every member of the biotic community’’ (1980, p. 327). These scholars argue, therefore, that bio-centric egalitarianism must be scrapped (Sylvan 1985). In a similar vein Fox has argued that the leveling axiology of orthodox Deep Ecology must be forsworn. If all organisms are really of equal intrinsic worth, the deep ecological doctrinaire might just as well eat veal as vegetables (Fox 1984). In reality, Fox predicted, deep ecologists probably tend to be vegetarians, because—in the words of Alan Watts—''cows scream louder than carrots'' (Fox 1984, p. 198).

Not eager to be labeled a procrustean ethicist, Fox persuasively argues for a position that abandons biocentric egalitarianism and instead asserts that all biota have intrinsic value but are not equal in intrinsic value because the ''richness of experience'' differs (Fox 1984, p. 198). On this point Fox aligns himself with the Whiteheadian-inspired environmental ethics based on intensity of sentience(Ferre´ 1994) that Sessions so adamantly opposes. To mark the difference between his sophisticated reformulation of deep ecological thinking from orthodox Deep Ecology, Fox rechristened his theory transpersonal ecology (1995). Fox has since moved beyond Deep Ecology and has developed a more integrated approach that encompasses inter-human ethics, the ethics of the natural environment, and the ethics of the human-constructed environment (Fox 2006).

In contrast, Sessions has reasserted the importance of deep ecology's ecological realism as opposed to social constructivism (2006) as the philosophical foundation for a ''new environmentalism of the twenty-first century'' (1995). Naess has steadfastly resisted any gradations or differentiations of intrinsic value among organisms in light of such criticisms. Responding to Fox, Naess wrote that some intrinsic values may differ, but not the kind he talks about. He and Fox, said Naess, ''probably do not speak about the same intrinsic view'' (Naess 1984, p. 202). Naess has reiterated his intuition that ''living beings have a right, or an intrinsic or inherent value, or value in themselves, that is the same for all of them'' (Naess 1984, p. 202). As Naess conceded early on (1973), brute biospherical reality entails some forms of killing, exploitation, and suppression of other living beings; the aim is to do more good than harm, to respect on an equal basis the right of every life form to flourish (Naess 1984).

Nevertheless, some philosophers have found such a guideline essentially vacuous, like vowing honesty until lying is warranted (Sylvan 1985a), thus undermining the very foundation of the principle itself. If any realistic practice deals with few situations where biota may be valued equally, then the principle is empty. According to some critics, there are irresolvable structural tensions between biocentric egalitarianism and metaphysical holism in ecological value systems (Keller 1997). They argue that, in light of the real functions of living natural systems, it is impossible to even come close to affirming both the ability of all individuals to flourish to old age and the integrity and stability of ecosystems. The necessity of exterminating ungulates such as goats and pigs for the sake of the health of fragile tropical-island ecosystems is but one example. Regard for the health of whole ecosystems might, therefore, require treating individuals differently, because individuals of different species have unequal utility (or disutility) for wholes; if that were the case, then viewed from the standpoint of an entire ecosystem, biocentric egalitarianism and metaphysical holism might be mutually exclusive and inconsistent with each other to the extent that at least one would have to be abandoned—or perhaps both (Keller 1997).

DEEP ECOLOGY, SOCIAL ECOLOGY, AND ECOFEMINISM

Social Ecologists and eco-feminists have also formulated robust critiques of Deep Ecology. Social Ecologists, speaking as secular humanists of the European Enlightenment tradition, have excoriated biocentric egalitarianism as misanthropic. In particular Murray Bookchin criticized Deep Ecology for reducing humans from complex social beings to a simple species, a scourge that is ''overpopulating'' the planet and ''devouring'' its resources (1988, p. 13).

Bookchin argues that Deep Ecologists' ahistorical ''zoologization'' prevents them from seeing the real cultural causes of environmental problems (1988, p. 18). In the estimation of ecological feminists, the idea of selfrealization is patriarchal. The Australian philosopher Val Plumwood, for instance, argued that the notion of the expanded self results in ''boundary problems'' stemming from the impulse of subordination (Plumwood 1993, p. 178). There are serious conflicts of interest between constituent members of larger wholes, and, she has argued, expansionary selfhood does not adequately recognize the reality of these conflicts. In the political arena, she contends, the expansionary holist is forced into the arrogant position of implying that anyone in disagreement does not in fact understand what is in her or his own best interest.

Instead of approaching a situation of conflicting interests with a conciliatory attitude (e.g., ‘‘I realize your interests are different from my interests, so here we have a real conflict of interest that we need to resolve by compromise’’), the expansionary holist approaches the situation, tacitly or overtly, selfrighteously (e.g., ‘‘I know what your real interests are, and here we have a conflict because you don’t seem to understand what your own interests are—whereas I do, fortunately for you.’’) Ecofeminists suspect that self-realization is a front for an imperialistic philosophy of self, springing from ‘‘the same motive to control which runs a continuous thread through the history of patriarchy’’ (Salleh 1984, p. 344). Consider the activist John Seed. According to the ecofeminist critique, there is nothing to guarantee that the needs of the rainforest should govern those of Seed: Why should Seed’s needs not dictate the needs of the rainforest? (Plumwood 1993). Or why should the needs of unemployed loggers not trump the needs of Seed and the forest?

Even while consenting to some of the insights of deep ecological questioning, for the unemployed logger the need to feed and clothe her or his children might easily outweigh any concern for ecosystemic integrity and stability. Furthermore, some eco-feminists argue, affirming the ontological interconnectedness of all human and nonhuman organisms and the nonliving environment does not necessitate an embrace of the holism of self-realization. In an article that has become required reading for students of Deep Ecology, the Australian philosopher Richard Sylvan notes that: ‘‘Certainly, removing human apartheid and cutting back human supremacy are crucial in getting the deeper value theory going. But for this it is quite unnecessary to go the full metaphysical distance to extreme holism, to the shocker that there are no separate things in the world, no wilderness to traverse or for Muir to save. A much less drastic holism suffices for these purposes’’ (1985b, p. 10).

CONCLUSION

Taken together, these various critiques have contributed to a significant consensus that Deep Ecology has reached its logical conclusion and has exhausted itself (Fox 1995). For example, in the respected textbook Environmental Philosophy (Zimmerman et al. 2005), the section on Deep Ecology, which enjoyed a coveted place in the first three editions, was eliminated in the fourth. Compared to other prominent theories, Deep Ecology has not crystallized into a complete system.

For Rothenberg (1996), Deep Ecology is a set of prescient ‘‘hints’’ about the real relations of culture and nature. These hints are to environmental philosophy as a tree trunk is to roots and branches (Rothenberg 1987). Inverting the apron diagram, Rothenberg visualizes the platform of Deep Ecology as a tree, its conceptual roots deriving nourishment from various religious, aesthetic, and speculative soils and its branches reaching out into the world, enjoining various types of political action (1987).

Rothenberg’s ideas have stimulated new ways of thinking about the ways in which humans experience nature and about the limits of human language (1996). Deep Ecology is less a finished product than a continuing, impassioned plea for the development of ecosophies (roots and branches) that merge shared non-anthropocentric core principles (the trunk). At the same time it is clear that Deep Ecology has earned a permanent and well-deserved place in the history of environmental philosophy; that this outlook has generated an abundance of academic articles and books in the field of environmental philosophy is ample testimony to its enduring influence and importance.

<u>Unit-5</u>

1.5 Organismic and Holistic Explanations

The interpretation of the organism-environment relation is central to both biology and psychology but is here usually assumed to follow the post-Cartesian paradigm whereby the organism is regarded as the active agent and the environment as something that only supports the actions of the organism or provides the basic factors for its niche. Hence the organism is perceived as agent and dynamic carrier of the mind and the environment as "objective" background for its actions. Inspired by the significant results of natural sciences, such as Newtonian physics based on the concept of isolated bodies, scientific psychology started with the application of objective methodology to the study of human behaviour and mental activity. In biology, the concepts of selection, environmental fitness, and adaptation were introduced, which were usually understood in terms of organisms born into an objective environment that sets the basic constraints for its survival. In psychology, the study of mental activity was characterized by such concepts as sensation and idea.

However, the ontological status of these concepts was not specified, and so they were usually thought to exist in their own world with thing like Newtonian properties or were reduced to the actions of the body and especially of the brain, with the rest of world being regarded only as a source of stimuli. The mainstream of experimental research adopted a reflex model: a stimulus-reaction connection, treating organism and the environment as two separate though connected parts. The reflex approach, however, was already being criticized by the end the19th century—not in the sense that the new direction had been completely wrong but rather that it did not go far enough, leaving alive all theoretical problems inherent in the old concepts of sensation and idea. Dewey (1896) maintained that the reflex approach did not solve the basic dualism of the introspectionist psychology but only replaced it with a dualism of peripheral and central structures, such that the relation between soul and body was transformed to a relation between stimulus and reaction.

According to Dewey, the reflex could not adequately integrate experimental facts because a stimulus-reaction relation cannot be regarded as a unit in the analysis of behavior. Stimulus and reaction are not independent successive processes but a whole determining each other, a coordination that is more a circle than a reflex arc. Later, Dewey examined human action especially as a system that always involves both parts of the body and environment (Dewey, 1922) and replaced organism-environment interaction with the concept of "transaction," stressing the simultaneous role of organismic and environmental constituents in the realization of the behavior of the organisms. At the same time, in biology, Uexküll and Kriszat (1932) developed Kantian ideas about organism-world relations, stressing the connection between the organism and environment, and divided the latter into two parts: die Umgebungund die Umwelt. Die Umgebung refers to the objective nature and die Umwelt to experienced environmental parts involved in the "Funktionskreis" connecting the organism and environment. Here, the subject-object division, for example, isn't located between the organism and the environment but between the perceptual and effector worlds, with the subject and object being intrinsically connected. In physiology, too, these kinds of ideas were launched at the beginning of the 20th century. Although physiological processes were generally considered something happening inside the organism, Haldane (1929) pointed out that the respiratory system may be conceived as a system only if no border is set between the "inner" and "outer" air.

This led Haldane to conclude that even more generally organism and environment cannot be separated by any exact border but form a unity that cannot be studied separately: the environment is as much dependent on the organism as the organism on the environment. Similar ideas were presented by Sumner (1922), who maintained the distinction between the organism and environment was only a conventional and arbitrary one. In psychology, the Gestalt school devoted special attention to the role of the environment, not as a source of stimuli as such but as a basic living condition of the organisms. In Gestalt psychology organism and environment were regarded as one unitary system, animal-environment system (Koffka, 1935).In analogy with the considerations of Uexküll, Koffka defined three different environments: phenomenal, behavioural, and geographic. The first was the subjectively experienced environment, the second the environment used in the behaviour of the organism, and the third the "objective" environment. According to Gestalt psychologists, all behaviour is realized in the behavioural environment. Behaviour is not based on one-way influence from the environment, but both the organism and the environment change and therefore must be regarded as only one system. Behavior entails reorganization of the whole system. In the stimulus-reaction conception, according to Koffka, a separation of the organism from the environment is postulated, which leads to the conception that in behaviour only the organism is changing, the environment and its stimuli staying constant. However, when the monkey eats a banana, the banana disappears and the monkey becomes satisfied.

Unit-6

1.6 Environmentalism in Geography

- **Introduction: The Status of the 'Environment' in Geographical Explanations**

The immediate stimulus is the perception among a number of geographers across the subfields of their discipline that 'the environment' has in certain spheres been brought into arguments that attribute it with powerful and singular causal power; and, moreover, that these arguments have been associated with the discipline in ways that have various effects on the nature of geo-graphical explanations and their public prominence (Watson et al, 2010).

In recent years a certain type of determinist environmental thinking has emerged. It can be understood to be one strand in a broader discourse of what we can call academic 'environment talk' (which includes political ecology, environmental history, climatology, and many others). Yet this 'neo-environmental determinism' (Sluyter, this Forum) is characterized by an emphasis on the core explanatory power of non-human/non-animal components of the biophysical sphere in shaping human outcomes (in relation to development, disease, conflict, re-sponses to climate change, etc.). Whereas other forms of environmental talk, such as political ecology (e. g., Peet and Watts, 2004), highlight the contingent historically and geo-graphically specific cultural meanings and human engagements with environmental processes, neo-environmental determinism claims to discern invariable dynamics be-tween (certain kinds of) society and 'the' environment. These determinist frameworks differ from early twentieth-century versions of environmental determinism in a number of ways.

First, in intellectual terms, recent determinisms have emerged in the context of widespread knowledge of Darwinian evolution, atmospheric and climate science, 'new genetics', and detailed ecological and social knowledge, all of which create a more knowledge-rich starting point (while also providing the basis of robust critiques of these same environmental determinisms; see below).

Second, in terms of socio-political con-texts, current environmental determinisms reflect subtle arguments about 'cultural' determinism, rather than crass racial ideologies. However, as Felipe Fernández-Armesto makes clear (this Forum), the concept 'culture' can be used as if it were equivalent to a biologically determined entity, and he thereby reveals the biologists that underlie certain strands of neo-environmental determinist thinking.

Many human geographers have expressed a combination of scepticism and surprise at the apparently inexorable rise of such arguments. Johnston (2007) calls Jared Diamond – whose books have often provided a lightning rod for critique and debate – 'a late inter-loper ... [but] not [a geographer] really'.

- **The Foundations of Scientific Thinking: the Society–Nature Dichotomy**

Science has a well-defined though not always conscious philosophical view in the background of investigations concerning the relationships between humans and nature: the separation of society from nature. This dichotomy permits the simplification of complex systems and the research of subsystems (Harden, C.P. 2012). According to Judkins, G. et al. (2008, p. 19) "the separation of humankind from nature, and the search for determinism within this relationship, are mutually constitutive and appear to varying degrees during all moments of human-environment research". Investigation of human–environment relations is one of the most basic questions of humankind. We can use the classical phrase that already the ancient Greeks had tried to explain, but we can go back even further, to the old myths of creation too. However, it was only in the second half of the 19th century when this research issue was considered as part of the academic science of geography.

Zoltán Hajdú (2007) considers this theme in a wider context. He agrees that this is one of the fundamental questions of geography, and further on, he introduces the evolution of scientific thinking in connection with these relationships. Basically, there are two contrasting viewpoints in science that have accompanied human history. The first is that environment controls social processes (determinism), and the second is that society has its own laws and nature is only the frame of its activities (nihilism).

Ferenc Probáld (2012) pointed out that several variants of determinism can be distinguished according to the supposed range of environmentally controlled historical processes or social phenomena, the degree and (in)directness of environmental impact. In Hungary, environmental determinism is a common phrase, but nihilism or possibilism is less known either in general or in scientific literature. These three phrases express theoretical ideas in the scientifi c approach of human-environment research (Castree, N. 2011; Harden, C.P. 2012; Judkins, G. et al. 2008; Lewthwaite, G.E. 1966; Peet, R. 1985; Probáld, F. 1999). Postmodern geographical approaches use the contemporary philosophical thought, namely that the nature is a social construction. It also implies that its meaning is under permanent change depending on philosophical approaches and political aspects too (Demeritt, D. 2002). We agree that human-environment relations were differently evaluated from time to time; therefore, we present here these changes through time and space (Kőszegi et al, 2015).

- **The Aspects of the Newly Professionalising Geography**

In the second half of the 19th and early 20th centuries the evolution theory of Darwinism, the deductive research methods and the Newtonian causality largely affected the scientific thought (Grossman, L. 1977). For geography, which was on its way to become an academic discipline, it was a problematic question how to treat the place and role of humans within the great natural system of the Humboldtian synthesis. The research on the relations between humans and environment has resulted in the genesis of anthropogeography. From a historical perspective, Hajdú (2007) claims, that the research on human-environment relations was a basic topic in the forming geographical science. According to István Berényi (1997) this issue is connected to the classic (early) anthropogeography, which analysed the connections between humans and their natural and social surroundings. Tibor Mendöl (1999), who wrote the history of geography in the middle of the 20th century, gave a different interpretation.

According to him, in the end of the 19th century geography was the science of connections and causalities in general. Our research group accepts the opinion of Probáld (1999), who emphasised that the integrated analysis of spatial phenomena and the investigation of human environment relations are among the most important targets of geographical research. Discussing the role of scientific approaches and paradigms in geography, scientists who helped the process of paradigm formation should also be mentioned. The research on human-environment relations has appeared in geography from time to time, but the connecting paradigms had different stories through time and space.

At the end of the 19th century the German, French, British and American schools were equally engaged in human-environment research and anthropogeography. Four scientists got important positions at different universities at the same time, and they were interested in the research of society within the framework of the new science investigating the features of the Earth (Castree, N. 2011). The German Friedrich Ratzel, the French Paul Vidal de la Blache, the British Halford Mackinder and the American William Morris Davis determined the scientific thought of human environment relationships within the new academic discipline of geography. All of them accepted the concept of unifi ed geography that the research of nature and society is feasible within one discipline. With Mackinder's words, they believed, that geography can bridge the gap between physical and social sciences (Castree, N. 2011). According to Davis the research on the relation between the Earth and its inhabitants is the task of geography, this research issue separate geography from other sciences (Lewthwaite, G.R. 1966; Harden, C.P. 2012). Their thought was influenced by the evolutionary theory of Lamarck and Darwin (Livingstone, D.N. 2011). We can say that it was a compulsion to them to demonstrate the relationship between nature and society.

- **Ratzel, Determinism and Their Influence on Scientific Thought**

The result of the activity of Friedrich Ratzel from the German School was the determinist paradigm about human-environment relations, which dominated geographical thought for some decades. Due to his work this paradigm got scientific legitimacy, but later on it had a controversial career in history, not only in a scientific meaning. It has been transformed through time and space, but basically it remained the same. According to environmental determinism the environment, the nature controls human activities (Livingstone, D.N. 2011). As Hajdú (2007) commented environment determines the diverse development processes of society.

Nature is the independent variable, the cause, while the human evolution and its social features are dependent variables, the answers to the cause (Harden, C.P. 2012). Environmental determinism was not the product of academic geography, discoveries had already made it popular. This idea was propagated by several earlier writings, and especially the influence of climate on people was a popular theory (Livingstone, D.N. 2002). In the 18th century, philosophers of the Enlightenment already wrote about the connections between the climate and the cultures. Geographic discoveries found various cultures at diff erent latitudes, which were dissimilar from the European culture; therefore, the relation between climate and culture seemed quite obvious (Coombes, P. and Barber, K. 2005). Merely a modern science was necessary to legitimate this viewpoint. Geography became an academic discipline more or less the same time when western powers demanded the legitimacy of their colonial aspirations (Livingstone, D.N. 1992). At the end of the 19th and early 20th centuries this thought was undoubtedly connected to geography and the reason for this was the subject of geographical research. According to Livingstone determinism served as a perfect basis for academic geography to provide an appropriate framework for the research of society. Second, it gave the scientifi c justifi cation of colonial policies and so the spirit of the age made it successful (Livingstone, D.N. 2011).

Peet, R. (1985) confirms this view in his article about the social background of environmental determinism; according to him this idea was the entrance of geography to modern sciences. Darwinian thoughts in geography gave a scientifi c explanation to the question, why it is possible that certain nations are more successful than others in the struggle for world domination (Peet, R. 1985). Consequently, environmental determinism is basically Eurocentric. Even nowadays we can meet scientifi c works based on the premise, that the formation of European culture was connected to special environmental features, or certain environmental features made non-European nations less resistant mentally and/or physically (Blaut, J.M. 1999; Castree, N. 2011). Ratzel's works are deeply inspired by the evolutionary theory; he studied zoology, biology and anatomy in the 1860s (Peet, R. 1985). He was a professor in Munich and later in Leipzig in the 1880s, when power efforts of the united German Empire became stronger; his thoughts gave the legitimacy of these imperialistic desires.

According to Berényi (1992), in the works of Ratzel the physical environment determines the possibility of human activities, the spatial movement of people and their spatial distribution; therefore the development of a state is the function of the physical settings. Mendöl (1999) emphasised that Ratzel had not claimed that every social phenomenon can be explained by environmental reasons; he just wanted to point those social phenomena, which really reflect the impact of environmental factors. In Ratzelian thought the state is an organism under the rule of biological evolution, like every creature on Earth. Nations live on a given territory, which feed them; therefore, the need for a larger territory or living space (Lebensraum) is instinctively present in their thoughts (Anderson, J. 2009). Later on, the living space theory had become notorious and compromised due to the book of Adolf Hitler (Mein Kampf), the Nazi ideology and the events of the Second World War. It is one of the reasons why environmental determinism disappeared from scientific thought and geopolitics in the second half of the 20th century.

However, the influence of Ratzelian thoughts is far beyond German geography and geopolitics. In his study about the short history of the 20th century geography Probáld (1999) discussed the predominance of environmental determinism in American geography in the first part of the 20th century too, thanks to the works of Ellen Churchill Semple and Ellsworth Huntington. Semple was Ratzel's student in the 1890s in Berlin. Her oft en cited study was published in 1911 (Infl uences of Geographic Environment) and became very infl uential for decades in the United States (Peet, R. 1985; Harden, C.P. 2012). Sometimes, the work of Semple is mentioned as a separate geographical approach as environmentalism (Lewthwaite, G.R. 1966, Probáld, F. 1999). In her convincing theory Semple emphasised the vitalising connection between Earth and man. Man cannot be investigated scientifically without the Earth, therefore, the aim of geography is to investigate the influence of natural factors on historical events (Peet, R. 1985). She investigated the effects of environment on human mind; this had involved the demonstration of mental features of nations and races. The basic thought, that the cradle of mankind is the hot zone, but the temperate zone off ers the challenges and trigger higher-order development, had already appeared in Ratzel's works. However, Semple went further: she described with spectacular examples the direct relation between nature and cultures (Peet, R. 1985). As Pál Teleki (1917/1996), wrote in his seminal work, Huntington went as far as to claim that the rise of civilisations is possible only in a certain climatic type of the Earth. In the works of Semple, Huntington and their followers the environmental factors were "determinative causes of racial diff erences, cultural practices, moral values, ingenuity and the ultimate capabilities of any given population" (Judkins, G. et al. 2008. p. 20). Looking back, they are criticisable, because they drew consequences without well-documented causes and effects and without systematic research. They generated many stereotypes and legitimated racism too (Harden, C.P. 2012).

According to Peet, R. (1985) determinism was popular in the United States in the early 20th century, because this theory legitimated the declaration of the superiority of the American nation as well as their spatial expansion over the American continent. Ratzelian thoughts were echoed in Hungarian geography much later. Research on human-environment relationships appeared only in the early 20th century due to the works of Jenő Cholnoky, a prominent physical geographer and Géza Czirbusz known as the Hungarian apostle of anthropogeography. While Cholnoky considered humans as one of the natural factors, Czirbusz advanced humankind from nature and he emphasised that other, more important internal effects have a significant role in the life of society (Fodor, F. 2006). Czirbusz considered Ratzel's thoughts and determinism with criticism and he called this theory "geographical fatalism". Therefore, Hajdú (2007) regards Cholnoky a deterministic scientist, whereas he considers Czirbusz a possibilistic or even a nihilistic thinker.

The Ratzelian concept of natural barriers was an important argument in the Hungarian struggle for the revision of the borders set by the Treaty of Trianon (1920). Probáld (2012) emphasised the presence of determinism in Hungarian geography between the two world wars. He presented several examples to demonstrate that the works of geographers were diff erently aff ected by this idea. Only Ferenc Fodor formulated extremely deterministic thoughts in his late work, when he stated that all functions of the state are deeply rooted in the geographical features of its land. According to him, the character of the nations bears strong imprint even of environments their ancestors lived in many centuries ago.

Nevertheless, other Hungarian geographers, who investigated human-environment relations like Pál Teleki, Tibor Mendöl, Gyula Prinz and András Rónai were closer to possibilism and the French School. Environmental determinism provided a scientific basis for the early 20th century scientists, who studied human phenomena in a changing world (Frenkel, S. 1992, 1994). According to Harden "the concept of environmental determinism, like the theory of continental drift , provided a stepping stone forthe advancement of knowledge" (Harden, C.P. 2012, p. 740). Nevertheless, determinism got more and more critics within the scientific community from the 1920s that has led to a paradigm shift in geography aft er the Second World War. However, this over-simplifying theory had great popularity and it influenced political decisions until the fall of colonizer politics (Frenkel, S. 1992, 1994).

- **The Critic of Determinism**

the impact of Paul Vidal de la Blache and the French School The infl uence of evolutionary theory is noticeable in the works of Vidal de la Blache too (he used the expression 'struggle for existence'), but as a historian he was rather a social scientist. Vidal de la Blache accepted the thought of unifi ed geography; nature and society exist in one integrated system in his works, but he examined their relations from the side of the society. Teleki (1917/1996) quoted his thoughts about geography: according to him geography received many ideas from other disciplines, but equally offers them a lot, because geography has the possibility to consider things together, that were intimately joined by nature and to understand and to make understand the relations of phenomena, which are present in the whole nature including all of us, humans, and the different landscapes.

According to Vidal de la Blache humans have a relative autonomy from nature, people rate and use natural resources in different ways (Berényi, I. 1997). His students phasised the importance of free will: "man is free to pick and choose between the vast but varying range of possibilities presented by his environment" (Lewthwaite, G.R. 1966, p 3; Teleki, P. 1917/1996). As Probáld (1999) wrote, the natural features could not determine the events of history, but provide a more or less wide range of possibilities. The utilisation of these possibilities depends on the cultural or technical development of the society. In possibilistic thought the nature is an eff ective but not deterministic factor in the formation of diff erences between cultures. Environment gives possibilities to social activities. The humans as actors create their own culture and their environment through this (Anderson, J. 2009). The French Jean Brunhes, a student of Vidal de la Blache, emphasised that researchers must concentrate on interrelationships and not on unidirectional relations (Lewthwaite, G.R. 1966).

According to Brunhes, as humans become members of their community and accept their culture through socialisation, they exert an impact on nature too. They become factors aff ecting the environment, but there are many other factors infl uencing the nature, therefore, infl uencing humans too (Teleki, P.

1917/1996). This is the essence of the human-environment relationship. His way of thinking was free from overstatements as it is refl ected by his claim that every truth related to human-environment relations can be only approximate, and the overemphasis on precision leads to falsifi cation (Brunhes, J. 1913). Vidal de la Blache examined smaller spatial units as opposed to the expanding state territories of his age; many landscape monographs were created by him and his followers (Teleki, P. 1917/1996). He coined the term genres de vie (way of life) and he pointed out that spatial behaviour of human groups is primarily affected by cultural features. He did not draw general conclusions, instead he wanted to explore concrete relationships first. That is why he turned back to earlier data collection and classification methodology. He wanted to gather the characteristics of groups with certain ways of life. His research was rather descriptive focussing on the quantitative and qualitative categorisation of all features in a landscape (Berényi, I. 1992; Mendöl, T. 1999; Anderson, J. 2009).

According to Berényi (1997), the possibilism theory was the successor of determinism in time; the Ratzelian thought became an obsolete conception by the turn of the 20th century due to the clear and intense transformation of nature by the upturning manufacturing industry. In fact, Vidal de la Blache and Ratzel were active in almost the same time; therefore, it is more appropriate to say that these two viewpoints lived next to each other. The infl uence of the French School and Vidal de la Blache penetrated to other countries, too. The concept of synthetic geography of Teleki, the prominent Hungarian geographer of the interwar period was closely connected to this approach. He was enthusiastic about the ingenuity of landscape monographs, but he considered them methodologically primitive (Teleki, P. 1917/1996). According to him, the mission of geographical description is to introduce the characters of landscapes and the comparison of them, searching for typical diff erences and similarities (Teleki, P. 1917/1996).

Possibilism could be used as a kind of scientific support to Hungarian irredentist efforts. Zoltán Krasznai (2003) pointed out that using ideas of the French School in the Paris Peace Conference was a tactical step. According to their concept, the Carpathian Basin is a complex of landscapes, which complete each other (Győri, R. 2009). The monograph of the Carpathian Basin is the last product of this idea (Bulla, B. and Mendöl, T. 1947). The infl uence of the French School can be recognised in the theoretical studies of István Dékány and in the works of Tibor Mendöl, too (Hajdú, Z. 2007; Győri, R. 2009). The predominance of descriptive geography became more and more obvious internationally till the remarkable paradigm shift after the Second World War. In the 1920s, the scientific arguments against determinism in the American geography used the approach of possibilism. These arguments and the basis of the human ecological approach are connected to Carl Sauer (Williams, M. 1994; Judkins, G. et al. 2008). In addition to the importance of the free will, Sauer emphasised that nature off ers or limits certain possibilities, but does not determine the culture (Harden, C.P. 2012). He stated that the human behaviour is not dependent on environmental constrains or on logical necessity but rather on the conventions acquired in the culture.

His research methodology took into consideration the historical development and used inductive methods like Vidal de la Blache, he presented how the culture and the physical environment can be studied in an integrated framework and context. (Judkins, G. et al. 2008). Another similarity, that Sauer performed his research using small territorial units too. He called them cultural landscapes, emphasising that they are the results of the joint infl uence of culture and nature (Harden, C.P. 2012). In the American geography, the ecological views appeared in the 1920s starting mainly with the research of Sauer who worked with some anthropologists at Berkeley University. His follower, Harlan Barrows emphasised that human ecological research can provide the appropriate framework for the unifi ed geography by the exploration of relationships between humans and the environment (Grossman, L. 1977; Harden, C.P. 2012). The early ecological studies of Sauer and his school concentrated principally to the prints of the society recognisable in the cultural landscape (Grossman, L. 1977).

Humans were emancipated from the environment through work and production and the society depending on its degree of development dominates nature (Smith, N. 1990). Not only the investigation of human-environment relationships, but even the hard separation of humans and nature became the ideological basis of the Marxist-Leninist geography. Nevertheless, during everyday research practices these ideas were in the service of actual political reasons and they were interpreted as it was advantageous for decision makers. According to Engel-Di Mauro (2009), the strict catechism followed by geographers was similar to the parody of Marx's works (Engel-Di Mauro, S. 2009). The geographical investigations were under state control, in service of the planned economy. Physical and human geography were separated from each other, and the later was replaced by economic geography, which pointed out the research directions (Timár, J. 2009). Science must be useful for society and it must serve the resolution of tasks set by the state, therefore, only applied research was favoured in the Eastern Block (Vavilov, Sz.I. 1950). On the other hand, science in the communist era was based on positivism, searched for objective truth and believed that the world is knowable (Vavilov, Sz.I. 1950).

Eastern Block geographers were rather thinking in a system of geographical sciences because of specialisation processes dissecting geography (Radó, S. 1962). Research was structured into two almost completely distinct units: physical and economic geography. Physical geography investigated the scene of production, the natural environment; therefore, it prepared the study of economic geography. In the Stalinist era, the task of the Soviet science was the service of monumental plans, like industrialisation, military preparations or the notorious environment-transformations (Shaw, D.J.B. and Oldfield, J.D. 2007). Practically, it led to the most simplistic interpretation of human-environment relations: society stands above nature and society is able to form and to transform nature in any way according to its needs (Shaw, D.J.B. and Oldfield, J.D. 2007). As geography served the coloniser ambitions of the Western countries earlier, so was it used by the Soviet politics to support the actual nature-transforming state plans. It was a total compulsion for them, they did not have a choice; they had to serve the dictatorship.

According to Hajdú (1999), the science of geography acted in fact in the propaganda of the works, and not in the formulation of plans. In the 1970s the negative environmental effects of the grand plans became so obvious that it inspired scientists to reconsider human-environment relationships again. Regional landscape research reappeared and new investigations with more qualitative methods as well as research themes from other fi elds of human geography (not economy) could begin (Shaw, D.J.B. and Oldfield, J.D. 2007; Timár, J. 2009). These developments were also valid for Hungary, where the sovietisation of science and of geography took place at the end of the 1940s. The end of this era when most studies neglected the environment can be assigned to the study of György Enyedi (1972). He discussed how much environmental factors were ignored in the study of social development. His work indicated the rethinking of nature-society relations in the early 1970s.

- **The Changing Western geography and the Nature-Society Dichotomy**

The concept of paradigm shift can explain the ignoration of human-environment relationships in geography after the Second World War (Kuhn, T.S. 1984). In the Western world the quantitative revolution and the spatial science approach, which endured till the 1980s pushed human-nature relationships aside during the second half of the 20th century (Probáld, F. 1999).

First, the abovementioned paradigm shift occurred mainly aft er the Second World War thanks to the specialisation of geographical research. Development of geomorphology, climatology, economic geography and political geography in the early 20th century was the forerunner of this fragmentation. Specialisation of human geography isolated the environmental studies. Physical geography focussed on Earth surface phenomena, but not in a holistic manner, instead even research on the physical environment was distributed into several sub-disciplines (Castree, N. 2011).

Second, the connection of determinism with coloniser efforts, racist views and Nazi ideas made it undesirable to both policy and society. Because of its intolerable situation, scientists discarded environmental determinism in the Anglo-Saxon world. As we mentioned earlier, environmental determinism was the entrance of geography to modern sciences, therefore, geography had to be rebuilt from its basis. The breaking with determinism pushed into background all kinds of research related to human-environment relationships. But there remained a vacuum after it in geographical science, and similar unifying paradigms have never appeared again since that time that would promote the investigation of human-environment relationships (Guelke, L. 1989). Besides this, Probáld (1999) explains the decline of determinism by a change of the way of thinking which appeared in the developed countries in the 1950s and 1960s. This new way of thinking is based upon the absolute faith in technical development that cannot accept any controlling act of nature. He did not call it nihilism, but he considered it as a backlash to the earlier deterministic thought.

After the Second World War the role of science in society has changed radically and geography also had to adjust to it. The quantitative revolution in geographical science as well as the spatial science approach of the discipline further reduced the connections between physical and human geography (Guelke, L. 1989). In the second half of the 20th century the human-environment research was basically ignored in geography but continued in other disciplines. Historians of the French Annales School analysed the relationships of different societies and the space around them (Braudel, F. 1949; Chaunu, P. 1966). After the specialisation of ecological research, there appeared some topics, which promoted the investigation of human-environment relationships, like cultural ecology, human ecology or political ecology. Cultural ecology became significant among anthropologists after the Second World War. Besides the relationships between cultures they investigated also the relations between different cultures and their environments (Grossman, L. 1977).

The ecological idea enriched the works of archaeologists too (Renfrew, C. and Bahn, P. 1996). Thanks to possibilism, human-nature research was present in geography too, but in a changed form and not in the focus of scientific attention. The ecological research as we mentioned above has already appeared in the early 20th century in American geography. Cultural and political ecology appeared aft er the second half of the 20th century; they interpreted the causality between humans and their environment from both directions (Harden, C.P. 2012). This idea presumed the correlations between special environmental characters and cultural traditions (Judkins, G. et al. 2008). Cultural ecology became significant particularly in American geography in the 1960s due to the works of Julian Steward, Roy Rappaport and Clifford Geertz (Castree, N. 2011).

They investigated the adapting processes of humans to nature (Harden, C.P. 2012). They focused on the changing processes caused by human activities (e.g. the effect of soil erosion, burning and cutting of vegetation), and analysed mainly the local features of smaller communities (Grossman, L. 1977). Political ecology investigated how the political and economic structures explained the interaction between society and its environment (Harden, C.P. 2012). According to the approach of structuralism in political ecology, the society is the main determining factor through its institutions (Judkins, G. et al. 2008). As the structuralism appeared in geography, the models of ecosystems worked out by biologists came also into use. These models were a great leap forward, because ecosystem analysis provides a useful framework to the investigation of mutual human–environment interactions (Grossman, L. 1977). However, instead of ecosystem analyses, the investigation of spatiality, and spatial analysis became dominant in geography.

The building of models and macro-regional investigations became characteristic, therefore, the ecosystem analyses, which were used mainly in small scale research, were not adopted. Nonetheless, the ecological research gave dynamics to the study of man-environment relationships again. Instead of looking for simple casual relationships, it revealed the complexity of links between humans, society and environment (Harden, C.P. 2012). These investigations focused mainly on smaller communities and territorial units of developing countries during the second half of the 20th century (Harden, C.P. 2012).

- **New Approaches around the Millennium**

The real breakthrough in human-environment research ensued in the 1990s, when the idea that humans have an influence on recent climate change was accepted (Coombes, P. and Barber, K. 2005). The environmental protection movements appeared first in the United States in the 1960s and 1970s and gradually gained political support to study these questions and increased research activities in these fields (Harden, C.P. 2012). The global climate change discourse received geopolitical importance and turned the attention to the fragile relation of humans and their environment. The environmental problems emphasised by politicians and the need for solving these problems generated a social claim towards science to study these questions (Judkins, G. et al. 2008). In the 21st century, investigation of human-environment relationships have become more significant not only in geography but in other social sciences too. These emerging issues have constituted a real challenge for geography.

In the past decades, due to Holocene research and new scientific methods, the investigation of human impact on natural environment became an important topic (Builth, H. et al. 2008; Lépy, É. 2012). The climate change discourse raises again the question that environmental changes can radically transform the life of societies. Many studies indicated correlation between climate change and cultural disasters (Coombes, P. and Barber, K. 2005). These studies emphasise the need for understanding these effects in order to reduce, stop or reverse the undesired results (Harden, C.P. 2012). There is a peculiar chapter in human-environment research, the investigation of factors, which mean risks to human communities and society needs protection against them (Castree, N. 2011). In this viewpoint, the natural factors are the independent variables again; they influence the life of communities (Harden, C.P. 2012). These viewpoints also gained more importance as recent climate change became a favourite subject. While the once ruling paradigm of environmental determinism was expelled from geography, it appeared again and even flourished (!) in other disciplines (Hulme, M. 2011). Biologists, historians, anthropologists and economists also investigate the role of natural factors in social processes and ask even basic questions like why certain nations are richer than others (Schoenberger, E. 2001).

While some economists, historians and climatologists formulated extremely deterministic and sometimes absurd statements (e.g. Landes, D.S. 1998, Behringer, W. 2010), the mainstream geography consistently rejected every sign of environmentalism (e.g. Blaut, J.M. 1999; Judkins, G. et al. 2008; O'Keefe, P. et al. 2009). However, we emphasize that geography must react to these environmentalist thoughts, in some cases even by adopting some less strict forms of environmental determinism (Diamond, J. 1997; Radcliffe, S.A. 2010). If any connection can be observed between environmental change and subsequent cultural transformation, the geographical community is inclined to think about deterministic relations (Nunn, P.D. 2003). Since these investigations are connected mainly to Quaternary research, the neo-deterministic approach appears principally in the works of physical geographers (Nunn, P.D. 2003; O'Keefe, P. et al. 2009).

- **The Challenges of the Future**

After we get acquainted with several works about human-environment relations, we can support the significance of this topic in our discipline. It is one of the most basic questions in geography; it gave the basis of becoming an academic discipline, and it greatly influenced its dual character in the 19th century. The nature-society dichotomy resulted different approaches in different periods and places, and it accompanied the whole history of geography. Recently, research on human-environment relationships reappeared due to the discourse of global climate change. Many scientists have denoted the risks and unscientific nature of classical deterministic thought (e.g. Sluyter, A. 2003). However, the ecological approaches, used in anthropology and archaeology, give an alternative, which emphasises the active role of people reacting to climate changes being in a dynamic relation with their own environment – they form and transform it (Ericson, C.L. 1999).

At the turn of the millennium the scientific community takes steps for the integration of ecological approaches with an actor oriented viewpoint. They would like to understand how the individuals can manipulate their own situations in the ecological, structural and cultural framework, in which they live (Judkins, G. et al. 2008). Despite the popularity of deterministic approach between laics and politicians, the scientific community investigates the human-environment relationships rather from an ecological point of view (Builth, H. et al. 2008; Lépy, É. 2012; Raymond, C.M. et al. 2013). Postmodern, poststructuralist, and postcolonial approaches have radically changed the philosophy of human-environment research in social sciences. According to these viewpoints, every representation of the nature is a social construction, the manifestation of some kind of social power. Thus, these approaches turn the idea of environmental determinism inside out, and they also point to the fact that the mental separation of nature and society, which is the basic of most human-environment concepts, is a heritage of Western philosophy (Castree, N. 2011).

Not only the global problems or the changing ideology of postmodern world induce the science to investigate human-environment relations. Due to the information revolution more effective equipment and better analysis methods are available for the scientific community; therefore, it is worth rethinking the relationships between humans and their environment. The geographical science has to renew not only its philosophical basis and scientific terms, but the connections between the two sub-disciplines of geography (i.e. between physical and human geography) must be refreshed too. The scientific community frequently emphasises the importance of multidisciplinary research and in the case of geography, this multidisciplinary approach can be achieved by coordinating the physical and human geographical investigations. The success of this coordinated research can be a key factor in the survival or renaissance of our discipline. Human-environment studies may have an important contribution to these efforts. Perhaps it is time for geography to reconsider its suspiciousness and hypersensitivity against all variants of determinism.

1.7 Spaceship Earth

Spaceship Earth (or Spacecraft Earth or Spaceship Planet Earth) is a world view encouraging everyone on Earth to act as a harmonious crew working toward the greater good.

- **History**

The earliest known useis a passage in Henry George's best known work, Progress and Poverty (1879). From book IV, chapter 2:

> It is a well-provisioned ship, this on which we sail through space. If the bread and beef above decks seem to grow scarce, we but open a hatch and there is a new supply, of which before we never dreamed. And very great command over the services of others comes to those who as the hatches are opened are permitted to say, "This is mine!"

George Orwell later paraphrases Henry George in The Road to Wigan Pier:

The world is a raft sailing through space with, potentially, plenty of provisions for everybody; the idea that we must all cooperate and see to it that everyone does his fair share of the work and gets his fair share of the provisions seems so blatantly obvious that one would say that no one could possibly fail to accept it unless he had some corrupt motive for clinging to the present system.

In 1965, Adlai Stevenson made a famous speech to the UN, in which he said:

> We travel together, passengers on a little space ship, dependent on its vulnerable reserves of air and soil; all committed for our safety to its security and peace; preserved from annihilation only by the care, the work, and, I will say, the love we give our fragile craft. We cannot maintain it half fortunate, half miserable, half confident, half despairing, half slave—to the ancient enemies of man—half free in a liberation of resources undreamed of until this day. No craft, no crew can travel safely with such vast contradictions. On their resolution depends the survival of us all.

The following year, Spaceship Earth became the title of a book by a friend of Stevenson's, the internationally influential economist Barbara Ward.

Also in 1966, Kenneth E. Boulding, who was influenced by reading Henry George, used the phrase in the title of an essay, The Economics of the Coming Spaceship Earth. Boulding described the past open economy of apparently illimitable resources, which he said he was tempted to call the "cowboy economy", and continued: "The closed economy of the future might similarly be called the 'spaceman' economy, in which the earth has become a single spaceship, without unlimited reservoirs of anything, either for extraction or for pollution, and in which, therefore, man must find his place in a cyclical ecological system". (David Korten would take up the "cowboys in a spaceship" theme in his 1995 book When Corporations Rule the World.)

The phrase was also popularized by Buckminster Fuller, who published a book in 1968 under the title of Operating Manual for Spaceship Earth.[6] This quotation, referring to fossil fuels, reflects his approach:

> …we can make all of humanity successful through science's world-engulfing industrial evolution provided that we are not so foolish as to continue to exhaust in a split second of astronomical history the orderly energy savings of billions of years' energy conservation aboard our Spaceship Earth. These energy savings have been put into our Spaceship's life-regeneration-guaranteeing bank account for use only in self-starter functions.

United Nations Secretary-General U Thant spoke of Spaceship Earth on Earth Day March 21, 1971 at the ceremony of the ringing of the Japanese Peace Bell: "May there only be peaceful and cheerful Earth Days to come for our beautiful Spaceship Earth as it continues to spin and circle in frigid space with its warm and fragile cargo of animate life."

Spaceship Earth is the name given to the 50 m (160 ft) diameter geodesic sphere that greets visitors at the entrance of Walt Disney World's Epcot theme park. Housed within the sphere is a dark ride that serves to explore the history of communications and promote Epcot's founding principles, "[a] belief and pride in man's ability to shape a world that offers hope to people everywhere." A previous incarnation of the ride, narrated by actor Jeremy Irons and revised in 2008, was explicit in its message:

Like a grand and miraculous spaceship, our planet has sailed through the universe of time, and for a brief moment, we have been among its many passengers….We now have the ability and the responsibility to build new bridges of acceptance and co-operation between us, to create a better world for ourselves and our children as we continue our amazing journey aboard Spaceship Earth.

David Deutsch has pointed out that the picture of Earth as a friendly "spaceship" habitat is difficult to defend even in a metaphorical sense. The Earth environment is harsh and survival is a constant struggle for life, including for whole species trying to avoid extinction. Humans wouldn't be able to live in most of the areas where they are living now without the knowledge necessary to build life-support systems such as houses, heating, water supply, etc.

The term "Spaceship Earth" is frequently used on the labels of Emanuel Bronner's products to emphasize and promote his belief in the unity of humankind.

1.8 Stationary State Economy

John Stuart Mill was among the nineteenth century's greatest philosophers. Perhaps best known for his defense of moral utilitarianism and individual liberty, Mill was also a leading economic thinker of his day. His major work on economics, Principles of Political Economy, built upon the insights of Adam Smith, David Ricardo, and Thomas Malthus in an effort to better systematize the principles of laissez faire economics and explain its "progress in wealth." Mill also explained, like his predecessors, why laissez faire economics would eventually culminate in what was then called a "stationary state economy," a condition of economic stagnation whereby a society, having reached the physical limits of economic growth, would simply reproduce wealth by replacing worn- out goods, maintaining capital stocks, and carefully husbanding non-renewable resources. But unlike his predecessors who viewed a stationary state as a dismal condition, Mill welcomed it; for he thought it gave people a sufficient level of wealth to both free them from life's necessary but coarser toils and provide them the leisure to develop their mental and moral capacities – necessary conditions for a happier life.

In this way, Mill tied his positive assessment of a stationary state to his moral defense of liberty and utility maximization. This contrasts with contemporary views of a stationary state economy, which typically see it as a hindrance to wealth creation and thus a constraint on utility maximization. Moreover, well- functioning markets within which economies grow are seen as neutral arbiters of competing consumer choices, and therefore protective of individual liberty. As a result, contemporary economists offer a positive assessment of economic growth and market liberalization based on liberty and utility maximization, turning Mill's view on its head (Buckley, 2011).

These opposing positions result from disagreement on how best to understand utility and describe markets. By retracing Mill's thought and contrasting it with similar defenses of laissez faire economics, the place and purpose of a stationary state economy can be revived for today's reader who might otherwise dismiss it as opposed to one's self-interest and irrelevant to today's issues.

Unit-8

1.9 Environmental Balance and Environmental Degradation

Environmental degradation is the disintegration of the earth or deterioration of the environment through consumption of assets, for example, air, water and soil; the destruction of environments and the eradication of wildlife. It is characterized as any change or aggravation to nature's turf seen to be pernicious or undesirable. Ecological effect or degradation is created by the consolidation of an effectively substantial and expanding human populace, constantly expanding monetary development or per capita fortune and the application of asset exhausting and polluting technology. It occurs when earth's natural resources are depleted and environment is compromised in the form of extinction of species, pollution in air, water and soil, and rapid growth in population.

Environmental degradation is one of the largest threats that are being looked at in the world today. The United Nations International Strategy for Disaster Reduction characterizes environmental degradation as the lessening of the limit of the earth to meet social and environmental destinations, and needs. Environmental degradation can happen in a number of ways. At the point when environments are wrecked or common assets are exhausted, the environment is considered to be corrupted and harmed. There are a number of different techniques that are being used to prevent this, including environmental resource protection and general protection efforts.

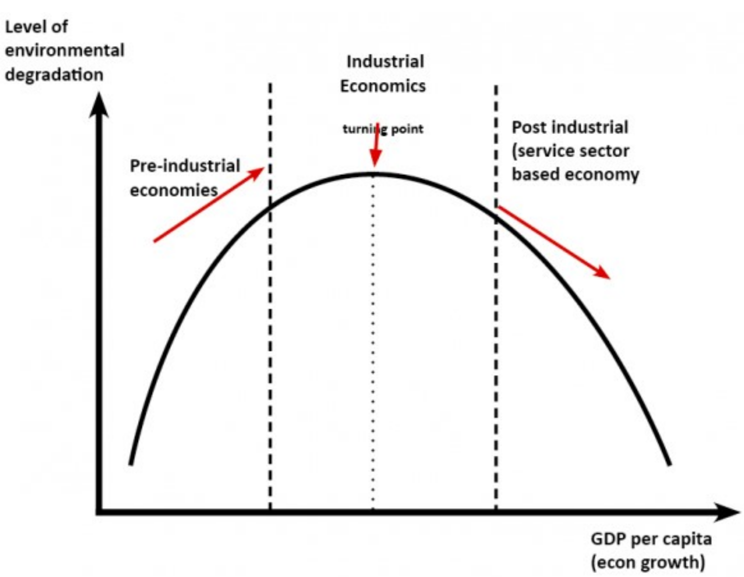
Understanding the Causes of Environmental Degradation

Economic growth seems to be the main driver of environmental degradation. We examine two countervailing factors in this section.

a) <u>Economic growth</u>

According to Grossman and Krueger (1995), economic development can affect environmental quality through the scale of economic activity, its composition (or structure) and the effect of income on the demand and supply of the pollution abatement effort. The larger the scale of economic activity, other things being equal, the higher the level of environmental degradation (pollution, resource depletion) is likely to be, since increased economic activity results in increased levels of resource use and waste generation.

The composition of economic activity affects environmental quality. Indeed, the primary sector tends to be more resource-intensive than the secondary or tertiary sectors. Industry (especially manufacturing), on the other hand, tends to be more pollution-intensive than either agriculture or services. Since the structure of the economy changes, part of the effect of income per capita reflects the effects of the changing composition of output. In other words, the composition effect can have a positive or negative impact on the environment because it measures the evolution of the economy towards a more or less appropriate productive structure. The technical effect is the positive environmental consequences of increases in income that call for cleaner production methods. Higher incomes enable higher public expenditure on environmental infrastructure as well as environmental regulations that drive private sector expenditure on abatement technologies. These three effects are illustrated by the Environmental Kuznets Curve (EKC). It assumes that environmental degradation increases up to a certain level of income; after this level, it decreases.



In addition, several authors consider that international trade may affect environmental quality through economic growth. Indeed, trade can influence environmental degradation through the scale, composition and technique effects. First, it raises economic activities that increase natural resource extraction and pollution (scale effect). It may change the type of economic activities to either less or more polluting industries (composition effect). Third, trade openness may encourage environmental production techniques. The EKC has been found for some pollutants and rejected for others. Indeed, Bimonte (2002) and Grossman and Krueger (1995) find an EKC for pollutants such as sulphur dioxide, carbon monoxide or nitrogen oxides. For carbon dioxide emissions, the hypothesis is rejected by HoltzEakin and Selden (1995). It is highlighted by Carson (2010), who shows that the corroboration of an EKC depends on econometric techniques, the quality of the data and the inclusion of other variables (Dinda 2004; Stern 2003; Stern 2004). Moreover, the EKC is a reduced form and does not shed light on the channels of transmission from economic activity to environmental degradation. Figure 1 below is an attempt to sort out several linkages that will be explored within this thesis. It illustrates for instance how environmental degradation may be dampened. This point is developed in the following section.

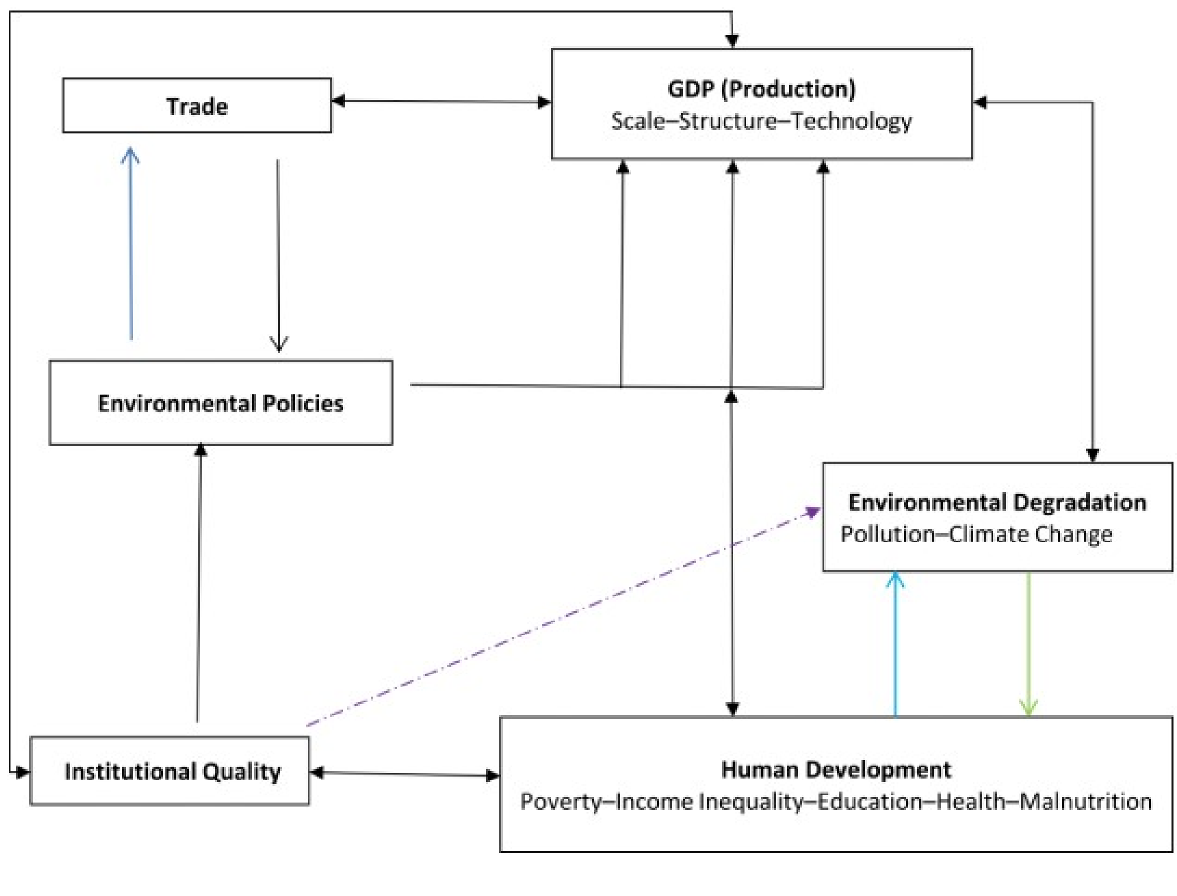


Figure: Relationship between Environmental Degradation and Economic Development

Source: Kinda, 2013: 6

b) Agricultural Runoff

Farming creates agriculture runoff issues. Agricultural runoff is a deadly source of pollutants which can degrade environments, so much so that the EPA identifies agriculture as the primary source of water pollution. Surface water washes over the soil and into lakes and streams. When it does so, it carries the fertilizers and pesticides used on the farm lands into water resources. Introducing poisons into waterways will have dire consequences. Fertilizers, whether or not they are organic, carry equal risks. Fertilizers containing large amounts of phosphorus can cause explosions of algae in lakes. As the algae die, bacteria start to breakdown the organic material. It soon develops into a situation where bacteria are using up the available dissolved oxygen in the water. Plants, fish, and other organisms begin to die off. The water becomes acidic. Like acid rain, lakes become dead zones with conditions so toxic that neither plants nor animals can live in these environments.

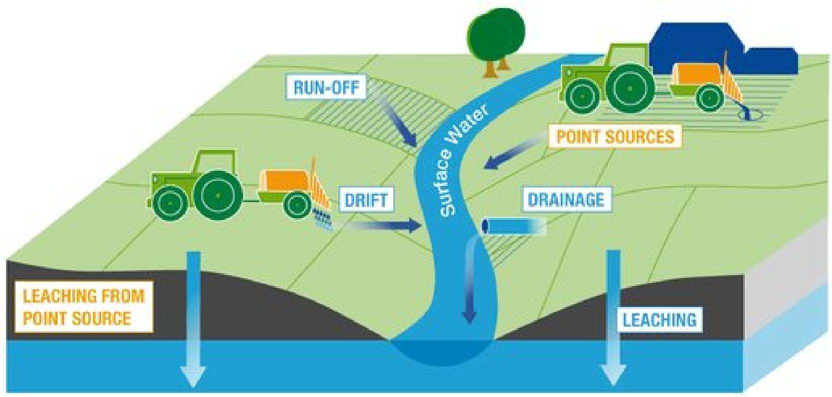


Figure: Agricultural Runoff and Water Pollution

SOCIAL FACTORS

c) Population

The rapid population growth and economic development in country are degrading the environment through the uncontrolled growth of urbanization and industrialization, expansion and intensification of agriculture and the destruction of natural habitats. One of the significant reasons for environmental degradation in India could be ascribed to quick development of population which is antagonistically influencing the natural resources and condition. The developing population and the ecological weakening face the test of maintained improvement without natural harm. The presence or the non-attendance of ideal characteristic assets can encourage or hinder the procedure of economic development. Population is an important source of development, yet it is a major source of environmental degradation when it exceeds the thresh hold limits of the support systems. Unless the connection between the multiplying population and the existence emotionally supportive network can be settled, improvement programs, howsoever, imaginative are not prone to yield wanted outcomes. Population impacts on the environment primarily through the use of natural resources and production of wastes and is associated with environmental stresses like loss of biodiversity, air and water pollution and increased pressure on arable land.

The increase in population has been due to the improvement in health conditions and control of diseases. The density of population has gone up from 117 in 1951 to 312 in 2001 and further to 382 persons in 2011 per square kilometre. A few push and draw factors are ventured to be agent towards trouble out relocation from rural to urban regions. This may be because of the declining asset accessibility per capita and contracting financial open doors in rural territories and better monetary openings, wellbeing and instructive offices and so on in urban regions giving chances to more elevated amount of human capital improvement could be the basic variables for country out movement. India supports 17 per cent of the world population on just 2.4 per cent of world land area.

d) Poverty

Poverty is said to be both cause and effect of environmental degradation. The round connection amongst poverty and environment is a to a great degree complex marvel. Imbalance may cultivate unsustainability in light of the fact that poor people, who depend on normal assets more than the rich, drain characteristic assets quicker as they have no genuine prospects of accessing different kinds of assets. As the 21st century starts, developing number of individuals and rising levels of utilization per capita are draining regular assets and corrupting the earth. The poverty-environmental damage nexus in India must be seen in the context of population growth as well. The pressures on the environment intensify every day as the population grows. The fast increment of human numbers joins with urgent poverty and rising levels of utilization are draining natural resources on which the vocation of present and future ages depends.

Poverty is amongst the consequences of population growth and its life style play major role in depleting the environment either its fuel demands for cooking or for earning livelihood for their survival. The unequal dispersion of assets and constrained open doors cause push and force factor for individuals living underneath poverty line that results in overburdened the population thickness in urban zones and condition get controlled by manifolds, subsequently, urban ghettos are produced in urban zones. Moreover, degraded environment can accelerate the process of impoverishment, again because the poor depend directly on natural assets. Although there has been a significant drop in the poverty ratio in the country from 55 per cent in 1973 to 36 per cent in 1993-94 and further to 27.5 per cent in 2004-05. The absolute number of poor has also declined from 320 million in 1993-94 to 301 million in 2004-05.

e) Urbanization

Urbanization in India started to quicken after freedom because of the nation's reception of a blended economy which offered ascend to the advancement of the private area. Urbanization is occurring at a quicker rate in India. Population living in urban territories in India, as per 1901 statistics, was 11.4%. This tally expanded to 28.53% as indicated by 2001 enumeration, and intersection 30% according to 2011 evaluation, remaining at 31.16%. As indicated by a review by UN State of the World Population report in 2007, by 2030, 40.76% of nation's population is required to dwell in urban zones. According to World Bank, India, alongside China, Indonesia, Nigeria, and the United States, will lead the world's urban population surge by 2050. Lack of opportunities for gainful employment in villages and the ecological stresses is leading to an ever increasing movement of poor families to towns. Such fast and spontaneous extension of urban areas has brought about debasement of urban condition. It has extended the hole amongst request and supply of infrastructural administrations, for example, vitality, lodging, transport, correspondence, instruction, water supply and sewerage and recreational pleasantries, along these lines exhausting the valuable ecological asset base of the urban areas.

f) Institutional Factors

The Ministry of Environment & Forests (MOEF) in the Government is responsible for protection, conservation and development of environment. The Ministry works in close coordinated effort with different Ministries, State Governments, Pollution Control Boards and various logical and specialized establishments ,colleges, non-Governmental associations and so on. Environment (Protection) Act, 1986 is the key legislation governing environment management. Other important legislations in the area include the Forest (Conservation) Act, 1980and the Wildlife (Protection) Act, 1972. The shortcoming of the current framework lies in the implementation abilities of natural foundations, both at the middle and the state. There is no effective coordination amongst various Ministries/Institutions regarding integration of environmental concerns at the inception/planning stage of the project. Current policies are also fragmented across several Government agencies with differing policy mandates. Absence of prepared work force and far reaching database postpone numerous activities. The greater part of the State Government organizations are moderately little experiencing deficiency of specialized staff and assets. Although overall quality of Environmental Impact Assessment (EIA) studies and the effective implementation of the EIA process have improved over the years, institutional strengthening measures such a straining of key professionals and staffing with proper technical persons are needed to make the EIA procedure a more effective instrument for environment protection and sustainable development.

g) Natural Causes

Environmental degradation has also been caused by natural processes. Among them forest fire, high intensity rainstorms, landslides, earthquakes etc. are important. Natural forest fire is common in Darjeeling hills during March-April when hill slopes become very dry and strong wind blows. The coniferous forest of high altitude has been found especially sensitive for such degradation. Large tracts around Singalila National Park affected by natural forest fire during the end of 19'h century which still remain visible even after a century. Landslide and associated phenomena is one of the most pervasive of natural processes that undermine the socio-economic development of Darjeeling hill areas. With rapid modernization, the Darjeeling Himalaya at present is experiencing a phenomenal growth in population (456%) during last 100 years). To cater to such an overwhelming population, pressure on the land is ever increasing. Forests have gradually been eliminated, steep slopes, generally unsuitable for human habitations and arable use have already been occupied and as a result Darjeeling Himalaya has of late turned into a highly vulnerable region without paying any heed to its ecological imbalance.

Types of Environmental Degradation

- **Climate Change**

There is no doubt that the composition of the atmosphere and the Earth‘s climate have changed since the industrial revolution predominantly due to human activities, and it is inevitable that if those activities do not shift markedly, these changes will continue regionally and globally. The atmospheric concentration of carbon dioxide has increased by over 30% since the pre-industrial era primarily due to the combustion of fossil fuels and deforestation. Global mean surface temperature, which had been relatively stable for over 1000 years, has already increased by about 0.75oC since the pre-industrial era, and an additional 0.5^{o}C to 1.0^{o}C is inevitable due to past emissions. It is projected to increase by an additional 1.2-6.4^{o}C between 2000 and 2100, with land areas warming significantly more than the oceans and Arctic warming more than the tropics.

Precipitation is likely to increase at high and middle latitudes and in the tropics, but likely to decrease in the subtropical continents. At the same time, evaporation increases at all latitudes. Over continents water is likely to be more plentiful in those regions of the world that are already water-rich, increasing the rate of river discharge and the frequency of floods. On the other hand water stress will increase in the sub-tropics and other water-poor regions and seasons that are already relatively dry, increasing the frequency of drought. Therefore, it is quite likely that global warming magnifies the existing contrast between the water-rich and water-poor regions of the world. Observations suggest that the frequencies of both floods and droughts have been increasing as predicted by the climate models. The Earth‘s climate is projected to change at a faster rate than during the past century. This will likely adversely affect freshwater, food and fiber, natural ecosystems, coastal systems and low-lying areas, human health and social systems. The impacts of climate change are likely to be extensive and primarily negative, and to cut across many sectors. For example, throughout the world, biodiversity at the genetic, species and landscape level is being lost, and ecosystems and their services are being degraded.

Although climate change has been a relatively minor cause of the observed loss of biodiversity and degradation of ecosystems, it is projected to be a major threat in the coming decades. There is a limit on the amount of fossil fuel carbon that we can pour into the atmosphere as carbon dioxide without guaranteeing climatic consequences for future generations and nature that are tragic and immoral. Given the decadal time scale required to phase out existing fossil fuel energy infrastructure in favor of carbon-neutral and carbon-negative energies, it is clear that we will soon pass the limit on carbon emissions. The inertia of the climate system, which delays full climate response to human-made changes of atmospheric composition, is simultaneously our friend and foe. The delay allows moderate overshoot of the sustainable carbon load but also brings the danger of passing a point of no return that sets in motion a series of catastrophic events. These could include melting of the Greenland and West Antarctic ice sheets leading to a sea level rise of many meters; melting of permafrost leading to significant emissions of methane, a potent greenhouse gas; and disruption of the ocean conveyor belt leading to significant regional climate changes.

These impacts would largely be out of human control. The risks from unmanaged climate change, as well as loss of biodiversity, are immense and action is urgent. Global warming due to human-induced increases in carbon dioxide is essentially irreversible on timescales of at least a thousand years, mainly due to the storage of heat in the ocean. Hence, decisions about anthropogenic carbon dioxide emissions being made today will determine the climate of the coming millennium. Even if emissions were to stop entirely in the 21st century, sea level would continue to rise. The level of carbon dioxide reached in this century will determine whether low lying areas are inundated by ice mass losses from Greenland and Antarctica, even if it occurs slowly over many centuries, because the warming will persist. The world‘s current commitments to reduce emissions are consistent with at least a 3 degree C rise (50-50 chance) in temperature. Such a rise has not been seen on the planet for around 3 million years, much longer than Homo sapiens have existed.

There is even a serious risk of a 5 degrees C increase, to an average temperature not seen on the planet for 30 million years. This is a problem for risk management and public action on a great scale. The fundamental market failure is the unpriced ―externality‖ of the impact of emissions. Other crucial market failures exist including those associated with R&D and learning, networks/grids, information, and further market failures around co-benefits such as valuation of ecosystem services and biodiversity issues. Policy will fail to generate the scale and urgency of the response required if it considers only the emissions market failure. The global community‘s attempts to address climate change have been hopelessly inadequate. The costs of climate change, already projected at 5% or more of global GDP, could one day exceed global economic output if action is not taken. The globe requires bold global leadership in governments, politics, business and civil society to implement the solutions, which have been scientifically demonstrated and supported by public awareness, to save humanity from climate change catastrophe.

- **Biodiversity, Ecosystems and their Services**

Biodiversity – the variety of genes, populations, species, communities, ecosystems, and ecological processes that make up life on Earth – underpins ecosystem services, sustains humanity, is foundational to the resilience of life on Earth, and is integral to the fabric of all the world‘s cultures. Biodiversity provides a variety of ecosystem services that humankind relies on, including: provisioning (e.g. food, freshwater, wood and fiber, and fuel); regulating (e.g. of climate, flood, diseases); cultural (e.g. aesthetic, spiritual, educational, and recreational), and supporting (e.g. nutrient cycling, soil formation, and primary production). These ecosystem services contribute to human wellbeing, including our security, health, social relations, and freedom of choice and action, yet they are fragile and being diminished across the globe. We are at risk of losing much of biodiversity and the benefits it provides humanity. As humankind‘s footprint has swelled, unsustainable use of land, ocean, and freshwater resources has produced extraordinary global changes, from increased habitat loss and invasive species to anthropogenic pollution and climate change. Threats to terrestrial and aquatic biodiversity are diverse, persistent, and, in some cases, increasing.

The Millennium Ecosystem Assessment concluded that 15 of the 24 ecosystem services evaluated were in decline, 4 were improving, and 5 were improving in some regions of the world and in decline in other regions. Action is critical: without it, current high rates of species loss are projected to continue what is becoming the 6th mass extinction event in Earth‘s history. It has been estimated that for every 1oC increase in global mean surface temperature, up to 5oC, 10% of species are threatened with extinction. All species count, but some more than others at any given time and place. Losing one key species can have cascading effects on the delivery of ecosystem services. Ecosystem services are ubiquitous, benefiting people in a variety of socioeconomic conditions, across virtually every economic sector, and over a range of spatial scales, now and in the future. The benefits that ecosystems contribute to human well-being have historically been provided free of charge, and demand for them is increasing. Although the global economic value of ecosystem services may be difficult to measure, it almost certainly rivals or exceeds aggregate global gross domestic product, and ecosystem benefits frequently outweigh costs of their conservation. Yet environmental benefits are seldom considered in conventional economic decision-making, and costs and benefits often don‘t accrue to the same community, or at the same time or place.

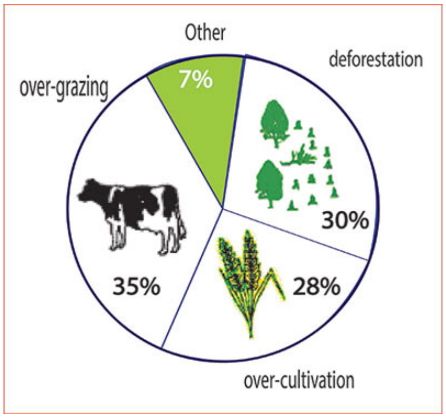
The value of these ecosystem services is being increasingly appreciated by a very large sector of society - extending from local stakeholders, the business community, agriculture, conservation, and governmental policy makers, including development agencies. Their economic value is enormous and a fundamental element of green economic development. However, we are degrading these services and squandering our natural capital for short-term gains. Two thirds of ecosystem services are currently being degraded globally, which will soon amount to an estimated loss of $500 billion annually in benefits. Green economic development will require technology development and technology transfer in order to increase value added from biological resources, especially in developing countries. This would help shift from the resource exploitative method of conventional development to the resource enrichment method of sustainable development.

- **Water Security**

Projections show that by 2025 over half of the world‘s population will live in places that are subject to severe water stress, and by 2040 demand is projected to exceed supply. This is irrespective of climate change, which will likely exacerbate the situation. Water quality is declining in many parts of the world, and 50 to 60% of wetlands have been lost. Human-induced climate change is projected to decrease water quality and availability in many arid- and semi-arid regions and increase the threats posed by floods and droughts in most parts of the world. This will have far-reaching implications, including for agriculture: 70% of all freshwater withdrawn from rivers and aquifers is currently used for irrigation. Of all irrigation water use 15 to 35% of irrigation water use already exceeds supply and is thus unsustainable. Freshwater availability is spatially variable and scarce, particularly in many regions of Africa and Asia.

Numerous dry regions, including many of the world‘s major ―food bowls,‖ will likely become much drier even under medium levels of climate change. Glacier melt, which provides water for many developing countries, will likely decrease over time and exacerbate problems of water shortage over the long term. Runoff will decrease in many places due to increased evapotranspiration. In contrast, more precipitation is likely to fall in many of the world's wetter regions. Developed regions and countries will also be affected. For example, Southern Europe in summer is likely to be hotter and drier.

- **Land Degradation**



Land degradation is any change or disturbance to the land perceived to be undesirable. Land degradation can be caused by both manmade and natural reasons such as floods and forest fires. It is estimated that up to 40 per cent of the world's agricultural land is seriously degraded. The main causes of the land degradation includes climate change, land clearance and deforestation, depletion of soil nutrients through poor farming practices, overgrazing and over grafting. In India, water erosion is the most prominent reason of land degradation. The growing trends of population and consequent demand for food, energy, and housing have considerably altered land-use practices and severely degraded India's environment. The growing population put immense pressure on land intensification at cost of forests and grazing lands because the demand of food could not increase substantially to population. Thus, horizontal extension of land has fewer scopes and relies mostly on vertical improvement that is supported by technical development in the field of agriculture i.e. HYV seeds, Fertilizers, Pesticides, Herbicides, and agricultural implements. All these practices are causing degradation and depletion of environment.

Figure: Causes of Land Degradation

- **Air pollution**

Air pollution in India is a serious issue with the major sources being fuel wood and biomass burning, fuel adulteration, vehicle emission and traffic congestion. Air pollution is also the main cause of the Asian brown cloud, which is causing the monsoon to be delayed. India is the world's largest consumer of fuel wood, agricultural waste and biomass for energy purposes. Traditional fuel (fuel wood, crop residue and dung cake) dominates domestic energy use in rural India and accounts for about 90 per cent of the total. In urban areas, this traditional fuel constitutes about 24 per cent of the total. Fuel wood, agro-waste and biomass cake burning releases over 165 million tons of combustion products into India's indoor and outdoor air every year. These biomass-based household stoves in India are also a leading source of greenhouse emissions contributing to climate change. On per capita basis, India is a small emitter of carbon dioxide greenhouse. In 2009, IEA estimates that it emitted about 1.4 tons of gas per person, in comparison to the United States' 17 tons per person, and a world average of 5.3 tons per person. However, India was the third largest emitter of total carbon dioxide in 2009 at 1.65 Gt per year, after China (6.9 Gt per year) and the United States (5.2 Gt per year). With 17 per cent of world population, India contributed some 5 per cent of human-sourced carbon dioxide emission; compared to China's 24 per cent share.

Effects of Environmental Degradation

1. **Impact on Human Health**: Human health might be at the receiving end as a result of the environmental degradation. Areas exposed to toxic air pollutants can cause respiratory problems like pneumonia and asthma. Millions of people are known to have died of due to indirect effects of air pollution.

2. **Loss of Biodiversity**: Biodiversity is important for maintaining balance of the ecosystem in the form of combating pollution, restoring nutrients, protecting water sources and stabilizing climate. Deforestation, global warming, overpopulation and pollution are few of the major causes for loss of biodiversity.

3**. Ozone Layer Depletion**: Ozone layer is responsible for protecting earth from harmful ultraviolet rays. The presence of chlorofluorocarbons, hydro chlorofluorocarbons in the atmosphere is causing the ozone layer to deplete. As it will deplete, it will emit harmful radiations back to the earth.

4. **Loss for Tourism Industry**: The deterioration of environment can be a huge setback for tourism industry that rely on tourists for their daily livelihood. Environmental damage in the form of loss of green cover, loss of biodiversity, huge landfills, increased air and water pollution can be a big turn off for most of the tourists.

5. **Economic Impact**: The huge cost that a country may have to borne due to environmental degradation can have big economic impact in terms of restoration of green cover, cleaning up of landfills and protection of endangered species. The economic impact can also be in terms of loss of tourism industry.

The Demand for Environmental Quality

It may be theoretically argued that the demand for environmental quality rises along with income (Vogel 1999). Indeed, economic growth is associated with increasing consumption and environmental degradation (waste and emissions). When people become rich and have higher living standards, their preferences for environmental protection increase. They may express their willingness to pay for environmental quality by supporting environmental policies in elections. The quality and the type of political institutions may allow people to express and support environmental policies.

a) Political institutions

A large body of literature has analysed the effect of political institutions on environmental quality. In cases in which environmental degradation has been mitigated, this may be attributed to local institutional reforms (Arrow et al. 1995). Payne (1995) argues that the members of the population in democratic countries are free to collect information about environmental degradation and are able to express their preferences and put pressure on their governments. With democracy, citizens are more aware of environmental problems (freedom of media). They can also express their preferences for the environment (freedom of expression) and create lobbying groups (freedom of association). Political leaders may become prompted (rights to vote) to implement environmental policies at the national and international levels. McCloskey (1983) and Payne (1995) highlight an important ability of democratic countries to satisfy people's environmental preferences and their willingness to commit themselves to international negotiations and agreements. Deacon (2009) and Olson (1993) argue that political freedoms favour environmental protection because non-democratic regimes will underproduce the environment considered as a public good. Autocratic governments are led by political elites who monopolize and hold a large share of the national incomes and revenues. The implementation of rigorous environmental policies can lower the levels of production, income and consumption, which, in turn, impose a higher cost on the elite in an autocracy than on the population, whereas the marginal benefit is uniform for both elite and population. Elites in an autocracy are therefore relatively less pro-environment than people in a democracy.

The empirical results are, however, not clear-cut. It may first be argued that democracy is positively linked with environmental commitment, but this is not necessarily the case with environmental outcomes (Neumayer 2002). Desai (1998) postulates that democracy does not protect the environment. Democracy is a factor of economic growth and prosperity, which may have a negative impact on the environment. Democracy is also correlated with factors such as property rights and social infrastructures that boost economic growth. First, Hardin (1968) worries about the management and overexploitation of natural resources. The property rights of environmental and natural resources (for example air, oceans, forests) are not well defined. This exploitation is accelerated in democracies where individuals have business and economic freedom. Moreover, General Introduction and Overview 8 Dryzek (1987) notices that democracies are also economic markets wherein lobbying groups are very important. Political leaders may be influenced by lobbying groups and multilateral companies. Democracies are not considered as protecting environmental quality as they are supposed to satisfy the preferences of markets and lobbying groups that aim to maximize their economic profit, which does not favour a better environmental quality.

b) The Role of Education

Educational attainment is generally considered a determinant of environmental preferences as well as an essential tool for environmental protection. Education enhances one's ability to receive, to decode and to understand information, and information processing and interpretation have an impact on learning and change behaviours (Nelson & Phelps 1966). In recent years (see Human Development Report (UNDP 2013); Global Environment Outlook-5 (UNEP 2012)), education has been considered as a vehicle for sustainable development. Some authors show that an increase in people's education is often accompanied by higher levels of environmental protection. Moreover, Farzin and Bond (2006) consider that educated people are more likely to generate an environmentally progressive civil service, and therefore have democratically minded public policymakers and organizations that are more receptive to public demands for environmental protection.

Unit-9

2.1 Man and Environment: Case Studies from River Valley Projects – Silent Valley and Narmada Dispute with Special Reference to Environmental Movement

Silent Valley Movement

"If the misery of our poor be caused not by the laws of nature, but by our institutions, great is our sin."
[Charles Darwin]

- **INTRODUCTION**

The mainstream development models implemented particularly in the Third World have had devastating effects on the ecology and environment of these countries to the extent of endangering the very existence of life forms on the planet earth. The basic problem with the development models prescribed and implemented is the unscrupulous exploitation of nature for short term economic gains of a few powerful individuals. The totalising imperatives of this new wave of development destroyed the forests and ultimately destabilised the '**blue planet'**. In the post-war period there was a strong faith in the development models proposed by the west. Most of the developing countries followed these dominant models without raising any question.

In India, the large dams were viewed as the "icons of development". Most often, the beneficiaries of these development models in India were the big farmers and the industrialists. Under these circumstances the traditional communities were totally marginalised and they were 'silenced' in the name of "nation building" and "development". In the process, they were evacuated from their homeland without any proper rehabilitation or compensation from the government. The most victimised among these sections are the *adivasis* and other deprived sections.

Large scale deforestation had taken place in the name of development in different parts of India. The destruction of the forests and the building up of huge dams resulted in the drying up of many water abundant rivers, other water bodies and ultimately deprived the local communities of their livelihood. The growing awareness among the *adivasis* and such excluded sections led to the emergence of large scale protest movements against the concerted efforts of the privileged sections of the society to conquer the forests in the name of development. The movement for protecting 'soil and water' has contributed to the contemporary discourse on development politics. There are assertions from the deprived sections against the well thought out designs of exploitation in the name of development and modernisation. People joined together cutting across parochial social and formal political divisions in the society to protect their right to livelihood. The civil society has become more active in environmental protection movements and gained acceptance beyond the formal political format. The strong feeling is that major political parties failed to address these types of issues. The need of the time is the active participation of the people in the forest conservation programmes and activities.

In this context we are discussing here the success story of an environmental protection movement in the Silent Valley in the Palghat district of Kerala against the proposed Hydroelectric project during the late 1980s. The movement was locally initiated with the strong support from the civil society and later on it was taken up by the Kerala Sastra Sahitya Parishad (KSSP) - the People's Science Movement of Kerala -and finally the government was forced to abandon the project. In a Post-Gandhian perspective, the movement was totally a non-violent one in a non-political space. As has already been noted, the modern development models failed to address the fundamental base of the human habitation, the environment. In this context, there are many attempts to go beyond the established models of development by local initiatives, which include the resistance against the annihilation of the forest reserve in the name of ‘development’ and also, the local initiatives for formulating a nature-friendly model of development. The movement in the Silent Valley is noted for its non-political character and admired for its seminal contribution towards creating awareness among the people on the importance of protecting the evergreen forests.

- **THE IMPORTANCE OF ‘SILENT VALLEY’**

The Silent Valley National Park is one of the last undisturbed rain forests and tropical moist evergreen forests in India. The park is located in the Nilgiri Hills, Palakkad District in Kerala, South India. The first English intrusion into the watersheds of the Silent Valley area was in 1847 by the botanist Robert Wight. The British named the area ‘Silent Valley’ because of a perceived absence of noisy Cicadas. It is estimated to have a continuous record of not less than 50 million years of evolution. In 1914 the forest of the Silent Valley area was declared a Reserve Forest. However, from 1927 to 1976 portions of the Silent Valley forest area were subjected to forestry operations.

The Silent Valley is rectangular, 7 km (east-west) X 12 km (north-south). Located between 11o03’ to 11o13’ N latitude and 76o21’ to 76o35’ E longitude, it is separated from the eastern and northern high altitude plateaus of the (Nilgiris Mountains) by high continuous ridges including Sispara Peak (2,206 m) at the north end of the park. The park gradually slopes southward down to the Palakkad plains and to the west it is bound by irregular ridges. The altitude of the park ranges from 658 m to 2328 m at Anginda Peak, but most of the park lies within the altitude range of 880 m to 1200 m. Soils are blackish and slightly acidic in the evergreen forests where there is good accumulation of organic matter.

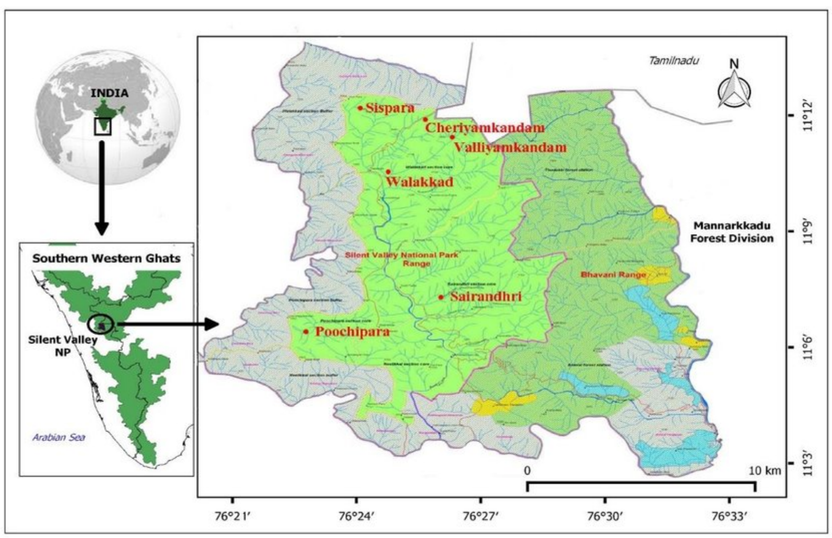


Figure: Location map of Silent Valley

The Kuntipuzha River stretches the entire 15 km length of the park from north to south and fall into the Bharathapuzha. Kuntipuzha divides the park into a narrow eastern sector of width 2 kilometers and a wide western sector of 5 kilometers. The river is characterised by its perennial crystal clear water. The main tributaries of the river, Kunthancholapuzha, Karingathodu, Madrimaranthodu, Valiaparathodu and Kummaathanthodu originate on the upper slopes of the eastern side of the valley. The river is uniformly shallow, with no flood plains or meanders. Its bed falls from 1,861 m to 900 m over a distance of 12 km, the last 8 km being particularly level with a fall of only 60 m. Kuntipuzha is one of the less torrential rivers of the Western Ghats, with a pesticide-free catchment area. Silent Valley gets copious amounts of rainfall during the monsoons, but the actual amount varies within the region due to the varied topography. The mean annual rainfall ranges from over 5000 mm in the Neelikal area in the west to around 3200 mm on the eastern side of the park. The park being completely enclosed within a ring of hills has its own micro-climate and probably receives some convectional rainfall in addition to rain from two monsoons. In general the rainfall is higher at higher altitude and decreases from the west to east due to the rain shadow effect.

Eighty per cent of the rainfall occurs during the south-west monsoon between June and September. It also receives significant amount of rainfall during the north-east monsoon between October and November. The mean annual temperature is 20.2°C. The hottest months are April and May when the mean temperature is 23°C and the coolest months are January and February when the mean temperature is 18°C. The relative humidity is consistently high (above 95%) between June and December because of the high rainfall. The valley is famous for many rare species of birds and animals. Birdlife International listed 16 bird species in Silent Valley as threatened or restricted: Nilgiri Wood-pigeon, Malabar Parakeet, Malabar Grey Hornbill, White-bellied Tree pie, Grey-headed Bulbul, Broad-tailed Grass bird, Rufous Babbler, Wynaad Laughing Thrush, Nilgiri Laughing Thrush, White-bellied Short wing, Black-and-rufous Flycatcher, Nilgiri Flycatcher, Whitebellied Blue-flycatcher, Crimson-backed Sunbird and Nilgiri pipit. Rare bird species found here include Ceylon Frogmouth and Great Indian Hornbill.

The 2006 winter bird survey discovered Long-legged Buzzard, a new species of raptor at Sispara, the park's highest peak. The survey found 10 endangered species recorded in the IUCN Red List including the Red winged crested cuckoo, Malabar Pied Hornbill, Pale harrier. The area is home to 15 endemic species including the Black-and-orange Flycatcher. It recorded 138 species of birds including 17 species that were newly observed in the Silent Valley area. The most abundant bird was the Black bulbul. The mammals in the valley include Gaur, largest of all wild cattle. There are at least 34 species of mammals at Silent Valley including the threatened Lion-tailed Macaque, Niligiri Langur, Malabar Giant Squirrel, Nilgiri Tahr, Peshwa's Bat (Myotis peshwa) and Hairywinged Bat. There are nine species of bats, rats and mice. Fourteen troops of lion-tailed macaque, eighty-five troops of Nilgiri langur, fifteen troops of bonnet macaque and seven troops of Hanuman langur were observed. Of these, the Nilgiri langur was randomly distributed, whereas the lion-tailed macaque troops were confined to the southern sector of the Park. Bonnet macaques and Hanuman langurs were occasional visitors.

The tiger, leopard (panther), leopard cat, jungle cat, fishing cat, Common Palm Civet, Small Indian Civet, Brown Palm Civet, Ruddy Mongoose, Stripe-necked Mongoose, Dhole, clawless otter, sloth bear, small Travancore flying squirrel, Indian pangolin (scaly anteater), porcupine, wild boar, sambar, spotted deer, barking deer, mouse deer and gaur also live here. There are at least 730 identified species of insects in the park. 33 species of crickets and grasshoppers have been recorded of which one was new. 39 species of true bugs (six new) and two species of Homoptera (both new) have been recorded. 128 species of beetles including 10 new species have been recorded. Over 128 species of butterflies and 400 species of moths live here. A 1993 study found butterflies belonging to 9 families. The families Nymphalide and Papilionidae contained the maximum number of species. 13 species were endemic to South India, including 5 species having protected status. 7 species of Butterflies were observed migrating in a mixed swarm of thousands of butterflies towards the Silent Valley National Park. In one instance an observer noted several birds attempting to catch these butterflies. The bird species included the Pied Bushchat Saxicola caprata, Nilgiri Pipit Anthus nilghiriensis, Tickell's Warbler Phylloscopus affinis, Greenish Leaf-Warbler Phylloscopus trochiloides and the Oriental White-eye Zosterops palpebrosa.

The flora of the valley include about a 1000 species of flowering plants, 108 species of orchids, 100 ferns and fern allies, 200 liverworts, 75 lichens and about 200 algae. In addition to facilitating recharge of the aquifer, water retention of the catchment basin and preventing soil erosion, every plant in the park from the smallest one celled algae to the largest tree in the forest has unknown potential for beneficial innovations in biotechnology. Angiosperm flora currently identified here includes 966 species belonging to 134 families and 599 genera. There are 701 Dicotyledons distributed among 113 families and 420 genera. There are 265 Monocotyledons here distributed among 21 families and 139 genera. Families best represented are the Orchids with 108 species including the rare, endemic and highly endangered orchids Ipsea malabarica, Bulbophyllum silentvalliensis and Eria tiagii, Grasses (56), Legumes (55), Rubiaceae (49) and Asters (45).

There are many rare, endemic and economically valuable species, such as cardamom Ellettaria cardamomum, black pepper Piper nigrum, yams Dioscorea spp., beans Phaseolus sp., a pest-resistant strain of rice Oryza Pittambi, and 110 plant species of importance in Ayurvedic medicine. Seven new plant species have been recorded from Silent Valley in 1996 including Impatiens sivarajanii, a new species of Balsaminaceae. Six distinct tree associations have been described in the valley. Three are restricted to the southern sector: (Cullenia exarillata & Palaquium ellipticum), (Palaquium ellipticum) and Mesua ferrea (Indian rose chestnut) and (Mesua ferrea & Calophyllum elatum). The remainders are confined to the central and northern parts of the Park: (Palaquium ellipticum & Poeciloneuron indicum), (Calophyllum elatum & Ochlandra sp.) and (Poeciloneuron indicum & Ochlandra sp.)

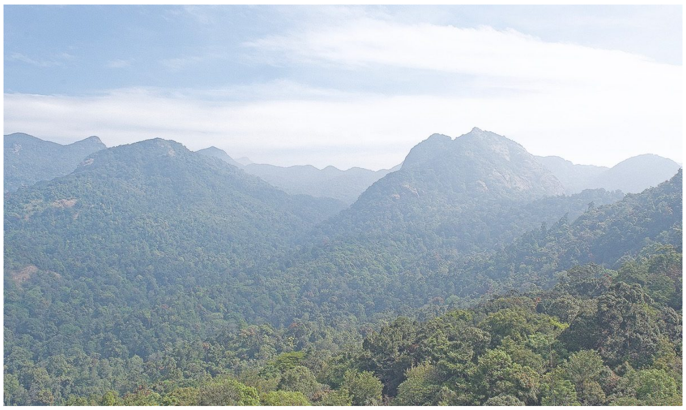
A study of natural regeneration of 12 important tree species of Silent Valley tropical rain forests showed good natural regeneration of all 12 species. The species studied were Palaquium ellipticum, Cullenia exarillata, Poeciloneuron indicum, Myristica dactyloides, Elaeocarpus glandulosus, Litsea floribunda, Mesua nagassarium, Cinnamomum malabatrum, Agrostistachys meeboldii, Calophyllum polyanthum, Garcinia morella and Actinodaphne campanulata. There is a huge hollow Kattualying tree here which can fit 12 people inside. Throughout human history about 10% of the genetic stock found in the wild has been bred into palatable and higher yielding cereals, fruits and vegetables. Future food security depends on the preservation of the remaining 90% of the stock through protection of high biodiversity habitats like Silent valley.

The National Bureau of Plant Genetic Resources of Indian Council of Agriculturaal Research, ICAR (India), Plant Exploration and Collection Division has identified Silent Valley as high in bio-diversity and an important Gene Pool resource for Recombinant DNA innovations. An important example of use of wild germplasm is gene selection from the wild varieties of rice Oryza nivara (Central India) and Oryza Pittambi found in Silent Valley for the traits of broad spectrum disease resistance in high yielding hybrid rice varieties including IR-36, which are responsible for much of the green revolution throughout Asia.

Plates: Biodiversity Hotspot of Silent Valley in Kerala

Small in Scale, Large in Impact

In terms of scale the Silent Valley is small. The area that was originally declared as a national park, and continues to be the core zone of the present demarcation, is 89.52 sq. km. The hydroelectric project that was envisaged in the valley was to have had an installed capacity of 120 megawatts. However, the controversy that the Silent Valley project generated in the 1970s and early 1980s made waves nationally and internationally. It was among the earliest significant milestones in the modern Indian environmental history.



"The people's movement against the Silent Valley project was significant in the sense that it was not a rhetorical protest, but one in which the scientific argument supported the views of the anti-dam group," Madhav Gadgil, the well-known ecologist who later went on to chair the Western Ghats Ecology Expert Panel (WGEEP), told Mongabay-India. Gadgil was a member of the multi-disciplinary committee constituted by the Government of India under the chairmanship of M.G.K. Menon, which finally proposed the cancellation of the Silent Valley project.

Gadgil commended the involvement of the Kerala Shastra Sahitya Parishad (KSSP), a people's science movement that was very popular in Kerala during the 1970s and 1980s, for this. "Theirs was a scientifically well-informed argument, and they could communicate it well with the public since they had the ability to reach out to the people in large numbers," he recalled.

According to environment journalist Darryl D'Monte in his book *Temples or Tombs* published by the Centre for Science and Environment in 1985, KSSP took a four-pronged approach in its communication. It emphasized that the Silent Valley was among the last such tropical evergreen forest patches remaining in the Western Ghats. Further, the organization stated the hydro-electric project was unjustified since 40% of Kerala's power generation was being exported to neighbouring Karnataka and Tamil Nadu; the project would contribute to only 7 per cent of the power produced in the state; and the 10,000 hectares proposed to be irrigated could be serviced through alternate ground and surface water at lower cost.

- **THE HYDRO ELECTRIC PROJECT**

At first, the Silent Valley got world attention not because of its rich natural diversity, but the successive struggle of the native people against the proposed hydroelectric project in the valley by the Kerala State Electricity Board [KSEB]. In 1928 the location on the Kunthipuzha River at Sairandhri was identified as an ideal site for electricity generation and in 1958 a study and survey of the area was conducted and a hydroelectric project of 120 MV costing Rs. 17 Crore was proposed by the Kerala State Electricity Board. Plans for a hydroelectric project that threatened the park's high diversity of wildlife stimulated an Environmentalist Social Movement in the 1970s called 'Save Silent Valley' which resulted in the cancellation of the project and the creation of the park in 1980. If the project was implemented, the reservoir would have submerged 8.3 km² of virgin rainforest and threatened the endangered Lion-tailed macaque. In 1976 the Kerala State Electricity Board announced a plan to begin dam construction and the issue was brought to the notice of the public.

The Kerala electricity board had started construction works in the proposed region. A vigorous public debate had taken place about the project. The scientific community, political parties, local people, activists and the civil society in general were included in this debate. However, the KSEB announced that it obtained clearance for the construction of the project from the Planning Board and the Science and Technology Department, Government of Kerala. But the fact is that it failed to obtain clearance from any of the concerned authority. The Morarji Desai government at the centre instructed the state government that sanction should be given to KSEB overruling the objections raised by the Science and Technology Department of Government of Kerala and the science community in the state. The project became a prestige issue for the KSEB, the Kerala government and the Prime Minister of the country. On the other side, the science community became more and more aware of the need for intensifying the resistance movement for protecting the rich diversity of the Silent Valley.

Finally the science community in Kerala understood that they could not, on their own, fight the cause and they felt that coordinated efforts of both the scientific community and the local people were inevitable for the conservation of the forests. It was very clear that the authorities who proposed the dam construction have not taken into consideration the kind of destruction that would set in motion in the entire geographical area. In this context it is worthwhile to note that the disastrous earthquake in Koyna in 1968 had been attributed to the weight of water in the Koyna reservoir. The effect of Aswan High Dam on the Nile Delta had been disastrous. Stoppage of flooding and consequent loss of fertility, increase of soil toxicity due to absence of flooding, reduction of still and consequent rapidity of current causing accelerated erosion of banks, stagnation of water and consequent increase in mosquitoes and diseases especially bilharzias had all added together. The Idukki Project Hydel project in Kerala submerged the entire natural forests. It was widely understood that the proposed dam would alter the eco-system of the valley.

- **MOVEMENT AGAINST THE PROJECT**

It is very important to note that an environmental movement like the Silent Valley Movement got national attention during the 1980s when the state-led development projects were dominating the scene. The dams were viewed as the 'icons' of development. The movement, in unequivocal terms, underlined the importance of protecting the environment for the generations to come. Protecting the lion tailed Macaque became the symbol of non–violent struggle to save the evergreen forests from total destruction. Many environmental groups like the Narmada Bachao Andolan (NBA), Bombay Natural History Society (BNHS) and Silent Valley Action Forum participated in the campaign. The prominent leaders were Vandana Shiva, Medha Patkar, Sundarlal Bahuguna, Baba Amte and Sunita Narain. The campaigns launched through the media, both print and electronic, generated public opinion in favour of the protection of the ecosystem of the silent valley.

As is stated elsewhere, the movement was first initiated by the local people and was subsequently taken over by the Kerala Sastra Sahitya Parishad (KSSP). Various scientific studies conducted by KSSP unequivocally emphasised the need for the protection of ecosystem in its pristine form for serving the interests of mankind. It was the biologist leaders of KSSP who identified the importance of protecting the valley because of its rich biodiversity. They had arranged several public meetings to educate the people. As a result most modern technical terms like the 'genetic diversity' became a household word in Kerala. Even concepts like 'gene pool' and 'deme' became part of the general vocabulary. The studies of the KSSP revealed that very limited section of the people would benefit from the proposed project; specifically, a large chunk of the electricity produced there would go for industrial purposes. Another argument was that by destroying the forest, the energy sources of vast majority of the poor people would be lost once and for all.

The KSSP generated public opinion against the project. It had science groups all over the state and through newsletters and journals it had spread the message among the students and youth as also the general public. It had sent a memorandum to the Kerala government about the issues and problems involved. It had organised street plays, exhibitions, public debates, and also conducted a 'marathon march' which covered around 400 villages. The student community also rose against the proposed project and it was the first time in the history of the state where the students agitated for the protection of the environment. Some celebrities who were actually not part of the environment movement like KPS Menon (Sr.) extended support to the cause of the Silent Valley. The campaign of KSSP was based on a distinct understanding of the following factors:

1. Vested interests, those who reap benefit from the felling of trees from deep forests and other similar activities, were getting protection from the powerful political class.

2. The achievements of Science and Technology are indiscriminately used for the promotion of the interests of the upper crest in the society.

3. It is the responsibility of the People's Science Movement to fight for the protection of the interests of common man and to be in the forefront of struggles to protect the environment to make the planet earth safe for the future generations.

4. Forest conservation is possible only through diverting people's struggles against the policies of the State, which is supporting the interests of certain sections of the population.

As a result of the massive campaign launched by KSSP, the Central Government appointed Dr. M.S. Swaminathan, a leading plant geneticist and agricultural Scientist to enquire about the issue. He visited the Silent Valley area and suggested that 389.52 km² including the Silent Valley (89.52 km²), New Amarambalam (80 km²), Attappadi (120 km²) in Kerala and Kunda in Tamilnadu (100 km²) reserve forests, should be developed into a National Rainforest Biosphere Reserve. In 1983, the Central Government instructed the State government to abandon the Project and on November 15, the Silent Valley forests were declared as a National Park. On September 7, 1985, the Silent Valley National Park was formally inaugurated. On September 1, 1986 Silent Valley National Park was designated as the core area of the Nilgiri Biosphere Reserve. Since then a long-term conservation effort has been undertaken to preserve the Silent Valley ecosystem. The Silent valley movement was in several ways crucial to other environmental movements in India. The seminal contribution of the Silent Valley movement is that it educated people regarding the importance of environmental protection. It is further realised that effective environmental protection can be achieved only through the active participation of local people in collaboration with the Science Community with the committed involvement of the civil society.

- **Prime minister takes a stand**

The decision to deny permission to the project is credited to Indira Gandhi. "Her role was indeed crucial as I have documented extensively in my book," said Jairam Ramesh, former environment minister and author of an ecological biography of Indira Gandhi published in 2017. "From 1979 onwards – influenced hugely by [ornithologist] Salim Ali and former diplomat K.P.S. Menon – she was against the project even though her own political colleagues in Delhi and Kerala were in support of the project." Ramesh's book gives instances of the prime minster corresponding with the chief minister of Kerala on protecting the forests.

According to Ramesh, the then agriculture secretary, M. S. Swaminathan also played an important role in convincing her about the need to abandon the project on ecological grounds. Gandhi encouraged debate and discussion on it for almost three years and set up a multi-disciplinary committee to assess the project.

"The Silent Valley represents one of the most coveted parts of our biodiversity heritage. Conserving this heritage was vital for ensuring the future of human wellbeing and happiness," Swaminathan said when asked about his recommendation to the government. He shared with Mongabay-India excerpts from a report he had written to the government after his tour of the valley in October 1978.

"The evolutionary age of the Silent Valley evergreen rainforest is believed to be more than 50 million years," Swaminathan wrote in his report. "The flora and fauna of this area are quite unique and 23 mammalian species including three endangered species – the tiger, lion-tailed macaque and the Nilgiri langur – have been recorded." He brought to the notice of the government a report by the Kerala Forest Research Institute that the forest tract was unique and relatively undisturbed.

"I was well aware during my years in the latter half of 1970s that our country's forests were being decimated by the construction of dams. Among these, the Silent Valley in Kerala, containing the richest tropical rainforest in the southern Western Ghats, was a conspicuous target of dam-builders against which there was already in Kerala a considerable public resentment, unmatched, however, by official response," said N.D. Jayal, then joint secretary for forests and wildlife in the Union Agriculture Ministry (the Ministry of Environment and Forests had not yet been formed), recalling how the events unfolded within the government.

All through her tenure as prime minister, Indira Gandhi had one nodal official who coordinated matters related to environment and forests. Jayal executed this role in the late 1970s and early 1980s.

"I found one day a file on my table seeking forest clearance for the Silent Valley project," Jayal said. "Feeling appalled, I took the file to consult the inspector general of forests who agreed that this was too precious a forest to be sacrificed. He said he felt helpless in these matters before the might of government, but agreed I could make a personal visit if there was a way to help save the forest."

In the Silent Valley, Jayal was "astounded by the sheer richness of the forest and its flora and fauna, worth saving at all cost." At Thiruvananthapuram, a remarkable young couple with PhDs in ecology – Sathis Chandran Nair and Santhi Nair – asked if they could show him a video presentation of the Silent Valley, of which they seemed to know every bit.

"The Nairs gave me a copy of their rich presentation, probably unavailable anywhere. I knew it would cut no ice with the government, and decided to show it and explain it to Salim Ali," Jayal told Mongabay-India. "I believe he also wrote about it to Prime Minister Indira Gandhi, who acted promptly by setting up the M.G.K. Menon committee, whose recommendation eventually led her to reject the dam project. However, we continued the fight until we were assured that the Silent Valley was declared a protected area as a national park."

Indira Gandhi decided to deny permission for the hydro-electric project on 18 October 1983, and it was notified as a national park on 15 November 1984, a fortnight after her assassination. Prime Minister Rajiv Gandhi inaugurated the park on 7 September 1985. In November 2009, while Jairam Ramesh was the environment minister, the Central and the Kerala Governments decided to add the adjoining forests of 147 sq.km as the buffer zone of the national park.

Many like Gadgil, who observed the events as they unfolded in those years, feel that Indira Gandhi's decision was motivated partially by her desire to conserve nature, and partially because of the international image that she had cultivated over the years as a head of state committed to the preservation of the environment.

- **Middle-class environmentalism?**

"The Silent Valley movement reached out to the educated and the middle-class intelligentsia," said Gadgil. "Earlier in the 1970s, there were wildlife conservation issues espoused by the former royalty. The Chipko (hug the trees) movement, on the other hand, was from the grassroots. The Silent Valley movement tapped into a different section of the society."

While this could have been one reason for the people's movement's impact, it also triggered the growth of environmental journalism in India. "It was an iconic issue," D'Monte told Mongabay-India. "Many environmental journalists cut their teeth on it and cited it for years afterwards. I am sure that is truer of Kerala, which has a more informed citizenry, as I point out in my book and, as a corollary, a more committed media. I can only presume that it is still referred to in Kerala in particular as well as Tamil Nadu." During the years of the controversy, D'Monte was editor of the Times of India and the Indian Express.

For many environmental journalists, the Silent Valley controversy continues to remain the point that triggered their interest in environment. "I was on a pilgrimage," wrote Bittu Sahgal, the editor of Sanctuary magazine when he visited the forest in 2012.

THE NARMADA BACHAO ANDOLAN

Water in a contemporaq society has undoubtedly become a site where contending perceptions meet. Recognizing the importance of the politics of water, Karl Wittfogel (1957) stated that every society is based on hydro-politics. The conflicts over water reflect the physical and symbolic character of water and can therefore focus on issues as diverse as particular physical infrastructure or non-state specific ideology. The water is requisite to human life inevitably means that water is also fundamental to culture. Therefore, systems must be in place to distribute water in ways that reduce the possibility of 'conflict. However, the contemporary societies have been experiencing a bourgeoning of water related conflicts throughout the world, especially with the rise and global spread of big dam building. Until 1960s-1970s, the water discourse of the multilateral agencies can be described as the ', hydraulic mission' (Swyngedouw: 1998; Reisner: 1984), commensurate with the drive towards modernity in which settlements and ecosystems submerged during construction of dams were seen as necessary sacrifices for the greater common good. Since then, Wade (1997) describes the upsurge of environmental projects, assessments and strategies within the World Bank (which is the chief funding agency for such projects) as 'mission overload', responding to the pressure from beyond areas immediately affected by Llams. Consequent to this, a negotiated 'complex multilateralism' emerges as the development agencies are influenced by, and form links with, social movements (0' Brien et al: 2000).

HISTORY AND EVOLUTION OF NARMADA BACHAO ANDOLAN

We all know that India got independence on August 15th, 1947. After getting independence India saw a new era emerging under the guidance and control of Jawaharlal Nehru who was decided to be the then Prime minister of India. Various investigations were carried out to get a conclusion on how to manage water from the Narmada river that flowed through the states of Gujrat, Madhya Pradesh, and Maharashtra and finally into the Arabian Sea. Disputes were going on between the states regarding the sharing of water. On October 6th, 1969 Narmada Water Disputes Tribunal was formed to adjudicate on the disputes that were going on. As per the orders of the Government of India, the tribunal investigated on the issues and after a long time of ten years responded back to the government with their valuable findings. As per the tribunal's decision, approval was given for the construction of 30 major, 135 medium, and 3000 small dams which also included the increase of the height of the Sardar Sarovar Dam. The decision was motivated by the assumption that by the end of the completion of all these projects forty million people will get better irrigation facilities and more electricity.

In 1985, after hearing about the Sardar Sarovar dam, Medha Patkar, and her colleagues, visited the construction area of the dam. After the visiting Medha Patkar and her colleagues got a clear picture of what was happening in the name of construction of the Sardar Sarovar Dam. People, mostly the Adivasis were said that they only need to relocate from where they presently are, and they will be given proper facilities in the place where they are relocated to, but the reality was something different. The people or whom we call *Adivasis* were asked to relocate or resettle even when the procedures of resettlement were not completed by the officials. The people initially relocated, even came back to the river banks as some of them got less fertile land or they were facing problems regarding fodder. And thus, on April 2nd, 1992 all these protests were formulized, and people together started to protest in Manibeli village, 8 kilometers away from the Sardar Sarovar Dam and they took an oath that even if the waters come after us we are ready to battle even against the water. Embarrassed with the people's attitude to drown rather than settle down, Operation Manibeli was organized. Hundreds of policemen were sent down to Manibeli village to move the people from the flood zone. With threats and giving more incentives, they managed to shift 13 families out of the flood zone. But the majority of the village vowed to stay back in the village.

FORMATION OF NARMADA BACHAO ANDOLAN

Narmada Bachao Andolan had various groups in it. Some of the prominent groups were Gujrat based Narmada Asargrastha Samithi (Committee of the affected people), Madhya Pradesh based Narmada Ghati Nav Nirman Samithi (Committee for a new life in Narmada Valley) and Maharashtra based Narmada Dharangrastha Samithi (Committee for Narmada dam-affected people). Narmada Bachao Andolan was also supported by various other NGO's and other famous personalities. NBA's slogans include - Vikas Chahiye, Vinash Nahin! (Development wanted, not destruction) and "Koi nahi hatega, bandh nahi banega!" (We won't move, the dam won't be constructed).

ROLE OF WORLD BANK

November 1992, World Bank President visits Bombay Louis Preston, the then World Bank president visited Bombay during that time. The World Bank has invested 450 million dollars in the Sardar Sarovar Dam. The anti-dam movement was rising to such an extent that pressure was too high on the World Bank. Thus, they finally decided to conduct a separate survey on the dam regarding the economic and technical terms. But the problem was that they forgot to analyze on the main concept that is social or environmental terms. Then also the committee found serious flaws in the project and recommended that the bank pull out. When Mr. Preston visited Bombay, Medha Patkar and some of her colleagues who are closely associated with the Narmada Bachao Andolan, decided to have a meeting with Mr. Louis Preston, but they were made disappointed.

Instead of him, he sent his assistant to talk with Medha Patkar and informed then that Mr. Preston has a busy schedule and is not able to meet them. Their main issue was that the Indian Government itself was against the *adivasis* as the government was funded by the World Bank itself. Therefore, these people wanted the World Bank to withdraw its funds. Later, the people found out that Mr. Louis Preston was engaged in attending a fashion show and he was least interested in talking to Medha Patkar and her associates. Outside the hotel, the policemen who were in charge took advantage of the absence of cameras and a brutal cane assault was done against the people and two of the activists were brutally murdered. After all the chaos, finally, the World Bank decided to withdraw from the project.

A Victory Rally was organized in Kapadia, Gujarat on March 1993 as the news of the withdrawal of the bank reached the people. Despite all the criticism, the dam was continued to be built. The drought-prone area of Kucch which was to be benefited from the project will receive only 2% of water; 30 years from now. Meanwhile, excess irrigation in the Command area will cause salinity and waterlog that will seriously affect the agriculture. Drastic deforestation led to siltation causing the lifespan of most of the dams to be reduced to small years. The siltation rate in the Narmada River is 365 times higher than the project was initially designed. We all know that electricity is the backbone of Industrial Revolution. Electricity is everything for us in today's life. But it is not like that for the Adivasis or the early men. For them, electricity is something that is irrelevant and unnecessary. According to them, electricity is needed by the government to run large factories. People can't drink or eat electricity and if you must access electricity then you must pay money. Electricity and the dam have destroyed the adivasis. But the State blames the adivasis. For building dams, the state gives the forest to big contractors to clear them down, but the state again blames the adivasis for cutting down the trees. But the adivasis use only a part of the forest. They can't cut down the whole of trees and transport it to anywhere because they lack such facilities. Before the coming up of Sardar Sarovar Dam, there were no roads in the hills. For the construction of roads, the officials were the ones who cleared the forests. Because the forests are cut down, hotness has increased rapidly and so diseases have started to spread a lot. It's considered a minimum sign of health when 33% of nation's land mass is covered by forests. Two decades ago, India's forest cover stood at 20%, today it is less than 10%.

MONSOONS 1993

DAM HEIGHT 61 CENTIMETERS

Monsoons were harsh. Monsoons were creating floods in the Narmada River. When we go deeper into the lives of the people living there, we can see that due to the flooding, police have forcefully evicted the people who were living there. They were forced to leave in the clothes they were wearing, and they couldn't even take any vessels or the things that they had. Most of the people took an oath that they will not relocate themselves even if the floods drown them. But the government thought that the people will fly off like rats when the real floods come. But their expectations were doomed. When the real floods came, people stood by their word and didn't move. But the government was not ready to accept their defeat. The policemen forcefully threw the people out of their homes during the night. It is like the people donated their homes to the government even though they didn't want it. Sulpaneshwar temple was drowned in the flood. It was the most important temple for the people living there. They still hoisted the flag over the top parts of the temple and showed their determination and strength.

Bombay, June 1993

The adivasis from the Narmada Valley marched in Bombay streets, claiming that the fodder, flowers, crops, cattle, wood, and grass all belong to them not the government. The village of Kutch is not getting any water. What will be the benefit of the project if the dam is going to give water to Kutch in 2025 and not now? People need water for their basic amenities and day to day need not only for future purposes.

And by the end of construction of the dam, from the fishermen to normal people will be affected to the core and who will be responsible for their loss; nobody, not even the Indian Government. This is not only the problem of the people living in the Manibeli village or the Narmada Valley, this is the problem of all the people who are oppressed and those who are not given their fundamental right to life and liberty. This fight will bring those who are still silent and will realize that they must protest and revolt if their rights are violated.

Right to life guaranteed by the Indian Constitution is equally applicable for all the people irrespective of their caste creed and color or the status in the society. As the officials acted as if they are least bothered about what is happening with the adivasis and the farmers, Medha Patkar and Deorambhai, decided to go on an indefinite hunger strike. On the 14th day, the hunger strikers were arrested, and the government promised that there will be a complete revision of the project and thus the hunger strike ended. A month later also no comprehensive steps were taken to revise the project and homes of 40 adivasi families were washed away. After realizing the fact that their hunger strike went in vain, several activists including Medha Patkar challenged the government that until and unless they bring in a revision to the plan they will drown themselves in the Narmada River on the 6th of August and titled the challenge as "Jal Samarpan" – Sacrifice by Drowning. The government invoked the "Official Secrets Act" restricting the media from the site. The police then launched a hunt for Medha Patkar and others. Thousands of the activists were arrested but the key activists avoided the police arrest.

5th August, 1993

When only 24 hours were only left for the Jal Samarpan, the Central Government agreed for a revision bit for the Gujarat ministers and businessmen the dam was bread and butter Manibeli, August 1993 Medha Patkar made the people understand that now they had to submit their facts and figures related to the Sardar Sarovar Dam in front of the Review Committee through her motivational and strong speeches. The villagers of Akhrani village resisted the Government's efforts to survey the land there suspecting that it may be a mere plan of the government to bring on evacuation. The police were so heartless that they fired on the people who oppressed the movement of the police.

2.2 Earth Summits: 1972 and 1992

The Stockholm and Rio Declarations are outputs of the first and second global environmental conferences, respectively, namely the United Nations Conference on the Human Environment in Stockholm, June 5-16, 1972, and the United Nations Conference on Environment and Development (UNCED) in Rio de Janeiro, June 3-14, 1992. Other policy or legal instruments that emerged from these conferences, such as the Action Plan for the Human Environment at Stockholm and Agenda 21 at Rio, are intimately linked to the two declarations, conceptually as well as politically. However, the declarations, in their own right, represent signal achievements. Adopted twenty years apart, they undeniably represent major milestones in the evolution of international environmental law, bracketing what has been called the "modern era" of international environmental law (Sand, pp. 33-35).

Stockholm represented a first taking stock of the global human impact on the environment, an attempt at forging a basic common outlook on how to address the challenge of preserving and enhancing the human environment. As a result, the Stockholm Declaration espouses mostly broad environmental policy goals and objectives rather than detailed normative positions. However, following Stockholm, global awareness of environmental issues increased dramatically, as did international environmental law-making proper. At the same time, the focus of international environmental activism progressively expanded beyond transboundary and global commons issues to media-specific and cross-sectoral regulation and the synthesizing of economic and development considerations in environmental decision-making. By the time of the Rio Conference, therefore, the task for the international community became one of systematizing and restating existing normative expectations regarding the environment, as well as of boldly positing the legal and political underpinnings of sustainable development. In this vein, UNCED was expected to craft an "Earth Charter", a solemn declaration on legal rights and obligations bearing on environment and development, in the mold of the United Nations General Assembly's 1982 World Charter for Nature (General Assembly resolution 37/7). Although the compromise text that emerged at Rio was not the lofty document originally envisaged, the Rio Declaration, which reaffirms and builds upon the Stockholm Declaration, has nevertheless proved to be a major environmental legal landmark.

- **Historical Background**

In 1968-69, by resolutions 2398 (XXIII) and 2581 (XXIV), the General Assembly decided to convene, in 1972, a global conference in Stockholm, whose principal purpose was "to serve as a practical means to encourage, and to provide guidelines … to protect and improve the human environment and to remedy and prevent its impairment" (General Assembly resolution 2581 (XXVI). One of the essential conference objectives thus was a declaration on the human environment, a "document of basic principles," whose basic idea originated with a proposal by the United Nations Educational, Scientific and Cultural Organization (UNESCO) that the conference draft a "Universal Declaration on the Protection and Preservation of the Human Environment".

Although there was general agreement that the declaration would not be couched in legally binding language, progress on the declaration was slow due to differences of opinion among States about the degree of specificity of the declaration's principles and guidelines, about whether the declaration would "recognize the fundamental need of the individual for a satisfactory environment" (A/CONF.48/C.9), or whether and how it would list general principles elaborating States' rights and obligations in respect of the environment. However, by January 1972, the working group managed to produce a draft Declaration, albeit one the group deemed in need of further work. The Preparatory Committee, however, loath to upset the compromise text's "delicate balance", refrained from any substantive review and forwarded the draft declaration consisting of a preamble and 23 principles to the Conference on the understanding that at Stockholm delegations would be free to reopen the text. At Stockholm, at the request of China, a special working group reviewed the text anew.

It reduced the text to 21 principles and drew up four new ones. In response to objections by Brazil, the working group deleted from the text, and referred to the General Assembly for further consideration, a draft principle on "prior information". The Conference's plenary in turn added to the declaration a provision on nuclear weapons as a new Principle 26. On 16 June 1972, the Conference adopted this document by acclamation and referred the text to the General Assembly. During the debates in the General Assembly's Second Committee, several countries voiced reservations about a number of provisions but did not fundamentally challenge the declaration itself. This was true also of the Union of Soviet Socialist Republics and its allies which had boycotted the Conference in Stockholm. In the end, the General Assembly "note[d] with satisfaction" the report of the Stockholm Conference, including the attached Declaration, by 112 votes to none, with 10 abstentions (General Assembly resolution 2994 (XXVII)). It also adopted resolution 2995 (XXVII) in which it affirmed implicitly a State's obligation to provide prior information to other States for the purpose of avoiding significant harm beyond national jurisdiction and control. In resolution 2996 (XXVII), finally, the General Assembly clarified that none of its resolutions adopted at this session could affect Principles 21 and 22 of the Declaration bearing on the international responsibility of States in regard to the environment. Following its adoption, in 1987, of the "Environmental Perspective to the Year 2000 and Beyond" (General Assembly resolution 42/186, Annex) – "a broad framework to guide national action and international co-operation [in respect of] environmentally sound development" - and responding to specific recommendations of the World Commission on Environment and Development (WCED), the General Assembly, by resolution 44/228 of 22 December 1989, decided to convene UNCED and launch its preparatory committee process.

The resolution specifically called upon the Conference to promote and further develop international environmental law, and to "examine … the feasibility of elaborating general rights and obligations of States, as appropriate, in the field of the environment". Work on this objective, and on "incorporating such principles in an appropriate instrument/charter/statement/declaration, taking due account of the conclusions of all the regional preparatory conferences" (A/46/48), was assigned to Working Group III (WG-III) on legal and institutional issues whose mandate was expanded beyond States' rights/obligations in the field of the environment, to include "development", as well as the rights/obligations of other stakeholders (such as individuals, groups, women in development, and indigenous peoples). WG-III held its first substantive meeting during the Preparatory Committee's third session in Geneva, in 1991. Actual drafting of the text of the proposed instrument, however, did not begin until the fourth and final meeting of the Preparatory Committee in New York, in March/April, 1992.

United Nations Audio-visual Library of International Law first advocated by a WCED legal expert group, did not win approval as it was specifically rejected by the Group of 77 developing countries (G-77 and China) as unbalanced, as emphasizing environment over development. The Working Group did settle instead on a format of a short declaration that would not connote a legally binding document. Still, negotiations on the text proved to be exceedingly difficult. Several weeks of the meeting were taken up by procedural maneuvering. In the end, a final text emerged only as a result of the forceful intervention of the chairman of the Preparatory Committee, Tommy Koh. The resulting document was referred to UNCED for further consideration and finalization as "the chairman's personal text". Despite threats by some countries to reopen the debate on the Declaration, the text as forwarded was adopted at Rio without change, although the United States (and others) offered interpretative statements thereby recording their "reservations" to, or views on, some of the Declaration's principles. In resolution 47/190 of 22 December 1992 the General Assembly endorsed the Rio Declaration and urged that necessary action be taken to provide effective follow-up. Since then, the Declaration, whose application at national, regional and international levels has been the subject of a specific, detailed review at the General Assembly's special session on Rio+5 in 1997, has served as a basic normative framework at subsequent global environmental gatherings, namely the World Summit on Sustainable Development in Johannesburg in 2002 and "Rio+20", the United Nations Conference on Sustainable Development in 2012.

Summary of Key Provisions and Their Present Legal Significance

a. General Observations

The Stockholm Declaration consists of a preamble featuring seven introductory proclamations and 26 principles; the Rio Declaration features a preamble and 27 principles. As diplomatic conference declarations, both instruments are formally not binding. However, both declarations include provisions which at the time of their adoption were either understood to already reflect customary international law or expected to shape future normative expectations. Moreover, the Rio Declaration, by expressly reaffirming and building upon the Stockholm Declaration, reinforces the normative significance of those concepts common to both instruments. Both declarations evince a strongly human-centric approach. Whereas Rio Principle 1 unabashedly posits "human beings … at the centre of concerns for sustainable development", the Stockholm Declaration — in Principles 1-2, 5 and several preambular paragraphs — postulates a corresponding instrumentalist approach to the environment.

The United Nations Millennium Declaration 2000 (General Assembly resolution 55/2), also reflects an anthropocentric perspective on respecting nature. However, the two declarations' emphasis contrasts with, e.g., the World Charter for Nature of 1982 (General Assembly resolution 37/7), and the Convention on Biological Diversity (preambular paragraph 1), whose principles of conservation are informed by the "intrinsic value" of every form of life regardless of its worth to human beings. Today, as our understanding of other life forms improves and scientists call for recognizing certain species, such as cetaceans, as deserving some of the same rights as humans, the two declarations' anthropocentric focus looks somewhat dated. At times Principle 1 of both the Stockholm and Rio Declarations has been mistaken to imply a "human right to the environment". The Stockholm formulation does indeed refer to a human's "fundamental right to … adequate conditions of life, in an environment of a quality that permits a life of dignity and well-being". However, at the conference, various proposals for a direct and thus unambiguous reference to an environmental human right were rejected. The Rio Declaration is even less suggestive of such a right as it merely stipulates that human beings "are entitled to a healthy. Since then, the idea of a generic human right to an adequate or healthy environment, while taking root in some regional human rights systems, has failed to garner general international support, let alone become enshrined in any global human rights treaty. Indeed, recognition of a human right to a healthy environment is fraught with "difficult questions" as a 2011 study by the United Nations High Commissioner on Human Rights wryly notes. As a basic UNCED theme, "sustainable development" — commonly understood as development that "meets the needs of the present without compromising the ability of future generations to meet their own needs"(Our Common Future) — runs like an unbroken thread through the Rio Declaration. However, sustainable development is also a strong undercurrent in the Stockholm Declaration, even though the WCED was not to coin the concept until several years after Stockholm. For example, Principles 1-4 acknowledge the need for restraint on natural resource use, consistent with the carrying capacity of the earth, for the benefit of present and future generations. The Rio Declaration expands on the sustainable development theme and significantly advances the concept by, as discussed below, laying down a host of relevant substantive and procedural environmental legal markers. Nevertheless, to this day the actual operationalization of the concept has remained a challenge. In this vein, on the eve of "Rio+20", United Nations Secretary-General Ban felt compelled to reiterate the urgent need for "sustainable development goals with clear and measurable targets and indicators."

b. The Prevention of Environmental Harm

Probably the most significant provision common to the two declarations relates to the prevention of environmental harm. In identical language, the second part of both Stockholm Principle 21 and Rio Principle 2 establishes a State's responsibility to ensure that activities within its activity or control do not cause damage to the environment of other States or to areas beyond national jurisdiction or control. This obligation is balanced by the declarations' recognition, in the first part of the respective principles, of a State's sovereign right to "exploit" its natural resources according to its "environmental" (Stockholm) and "environmental and developmental" policies (Rio). While at Stockholm some countries still questioned the customary legal nature of the obligation concerned, today there is no doubt that this obligation is part of general international law. Thus in its Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons first, and again more recently in the Case concerning Pulp Mills on the River Uruguay, the International Court of Justice expressly endorsed the obligation as a rule of international customary law. Moreover, the Pulp Mills decision clearly confirms that the State's obligation of prevention is one of due diligence.

c. The Right to Development in an Environmental Context

Both at Stockholm and at Rio, characterization of the relationship between environment and development was one of the most sensitive challenges facing the respective conference. Initial ecology-oriented drafts circulated by western industrialized countries failed to get traction as developing countries successfully reinserted a developmental perspective in the final versions of the two declarations. Thus, after affirming that "both aspects of man's environment, the natural and the man-made, are essential to his well-being" (preambular paragraph 1), Principle 8 of the Stockholm Declaration axiomatically labels "economic and social development" as essential. Rio Principle 3, using even stronger normative language, emphasizes that the "right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations". Although the United States joined the consensus on the Declaration, in a separate statement it reiterated its opposition to development as a right. The international legal status of the "right to development" has remained controversial even though, post-Rio, the concept has attracted significant support, e.g. through endorsements in the 1993 Vienna Declaration and Programme of Action, and the Millennium Declaration. At any rate, there is no denying that the Rio formulation has had a strong impact on the international political-legal discourse and is frequently invoked as a counterweight to environmental conservation and protection objectives. Today, economic development, social development and environmental protection are deemed the "interdependent and mutually reinforcing pillars" of sustainable development (Johannesburg Plan of Action, para.5). d. Precautionary Action One of several of the Rio Declaration Principles that does not have a counterpart in the Stockholm Declaration is Principle 15, which provides that "the precautionary approach shall be widely applied by States according to their capabilities:" Whenever there are threats of serious or irreversible damage, a lack of full scientific certainty shall not excuse States from taking cost-effective measures to prevent environmental degradation. At Rio, a European initiative proposing the inclusion of precautionary action as a "principle" failed to gain support. Today, the concept is widely reflected in international practice, although there exists no single authoritative definition of either its contents or scope. This has prompted some States, including the United States, to question its status as both a "principle of international law" and a fortiori a rule of customary international law (World Trade Organization, European Communities – Measures Affecting the Approval and Marketing of Biotech Products, paras.7.80-7.83). However, in its 2011 Advisory Opinion, the Seabed Chamber of the International Tribunal of the Law of the Sea takes note of "a trend towards making this approach part of customary international law", thereby lending its voice to a growing chorus that recognizes "precaution" as an established international legal principle, if not a rule of customary international law.

d. "**Common but Differentiated Responsibilities"**

While today the concept of "common but differentiated responsibilities" ("CBDR") is accepted as a cornerstone of the sustainable development paradigm, it is also one of the more challenging normative statements to be found in the Rio Declaration. The second sentence of Principle 7 provides: "In view of the different contributions to global environmental degradations, States have common but differentiated responsibilities". Ever since its adoption, its exact implications have been a matter of controversy. Specifically, taken at face value the formula seems to imply a causal relationship between environmental degradation and degree of responsibility. However, "differential responsibilities" has been considered also a function of "capability" reflective of a state's development status. Unlike the essentially contemporaneously drafted provision in the United Nations Framework Convention on Climate Change, which refers to States' "common but differentiated responsibilities and respective capabilities" (Article 3, para.1, emphasis added), the second sentence of Principle 7 omits any reference to capabilities. A separate sentence in Principle 7 does acknowledge the relevance of capabilities. But it does so in relation to developed countries' special responsibility regarding sustainable development on account of "the technologies and financial resources they command". Principle 7 indirectly, then, links developing country status to "responsibilities". What remains unclear, at any rate, is whether "CBRD" implies that developing country status in and of itself entails a potential diminution of environmental legal obligations beyond what a contextually determined due diligence standard would indicate as appropriate for the particular country concerned. Certainly, both the Stockholm and Rio Declarations (Principle 23 and Principle 11, respectively) expressly recognize the relevance of different national developmental and environmental contexts for environmental standards and policies purposes. However, developing country status per se does not warrant a lowering of normative expectations.

The United States stated for the record that it "does not accept any interpretation of Principle 7 that would imply a recognition or acceptance by the United States of … any diminution of the responsibilities of developing countries under international law". The United States delegation offered the same "clarification" in respect of various references to "CBDR" in the Plan of Implementation of the World Summit on Sustainable Development in 2002. Consistent with this view, the 2011 International Tribunal of the Law of the Sea Advisory Opinion, in construing the scope of a State's international environmental obligations, refused to ascribe a special legal significance to developing country status and instead affirmed that "what counts in a specific situation is the level of … capability available to a given State…".

e. **Procedural Safeguards**

Principles 13-15 and 17-18 of the Stockholm Declaration — rather modestly — emphasize the need for environmental and development planning. The absence of any reference in the Declaration to a State's duty to inform a potentially affected other state of a risk of significant transboundary environmental effects was due to the working group on the Declaration's inability to reach agreement on such a provision. However, the working group did agree on forwarding the matter to the General

Assembly which, as noted, endorsed such notification as part of States' duty to cooperate in the field of the environment.

By contrast, the Rio Declaration unequivocally and in mandatory language calls upon States to assess, and to inform and consult with potentially affected other States, whenever there is a risk of significantly harmful effects on the environment: Principle 17 calls for environmental impact assessment; Principle 18 for emergency notification and Principle 19 for (routine) notification and consultation. At the time of the Rio Conference, and perhaps for a short while thereafter, it might have been permissible to question whether the contents of all three principles corresponded to international customary legal obligations. However, today given a consistently supportive international practice and other evidence, including the International Law Commission's draft articles on Prevention of Transboundary Harm from Hazardous Activities, any such doubts would be misplaced.

g. Public Participation

Principle 10 of the Rio Declaration posits that "[e]nvironmental issues are best handled with the participation of all concerned citizens, at the relevant level". It then calls upon States to ensure that each individual has access to information, public participation in decision-making and justice in environmental matters. Although Principle 10 has some antecedents in, for example, the work of the Organization of Economic Co-operation and Development, it nevertheless represents a trail blazer, laying down for the first time, at a global level, a concept that is critical both to effective environmental management and democratic governance. Since then, international community expectations, as reflected notably in the Convention on Access to Information, Public Participation in DecisionMaking and Access to Justice in Environmental Matters (Aarhus Convention), the 2010 UNEP Guidelines for the Development of National Legislation on Access to Information, Public Participation and Access to Justice in Environmental Matters and various resolutions of international organizations and conferences, have coalesced to the point where the normative provisions of Principle 10 must be deemed legally binding. While the actual state of their realization domestically may be still be a matter of concern— implementation by States of their Principle 10 commitments is specifically being reviewed within the context of Rio+20—today the rights of access to information, public participation, and access to justice arguably represent established human rights.

h. The Interface of Trade And Environment

In Principle 12 of the Declaration, the Rio Conference sought to address one of the controversial issues of the day, the interrelationship between international trade and environmental conservation and protection. After exhorting States to avoid trade policy measures for environmental purposes as "a means of arbitrary or unjustifiable discrimination or a disguised restriction on international trade" — language that closely follows the chapeau of Article XX of the General Agreement on Tariffs and Trade (GATT) — Principle 12 criticizes States' extra-jurisdictional unilateral action: "Unilateral actions to deal with environmental challenges outside the jurisdiction of the importing country should be avoided". This provision traces its origin to a proposal by Mexico and the European Community both of which had been recent targets of United States environment-related trade measures. Responding to the adoption of Principle 12, the United States offered an interpretative statement that asserted that in certain circumstances trade measures could be effective and appropriate means of addressing environmental concerns outside national jurisdiction. This U.S. position has now been fully vindicated. As the World Trade Organization Appellate Body first acknowledged in the Shrimp-Turtle cases, unilateral trade measures to address extraterritorial environmental problems may indeed be a "common aspect" of measures in restraint of international trade exceptionally authorized by Article XX of the GATT.

i. Indigenous People

Rio Principle 22 emphasizes the "vital role of indigenous people and their communities and other local communities" in the conservation and sustainable management of the environment given their knowledge and traditional practices. It then recommends that States "recognize and duly support their identity, culture and interests and enable their effective participation in the achievement of sustainable development". Even at the time of its drafting this was a somewhat modest statement, considering that in the case of indigenous peoples, cultural identity and protection of the environment are inextricably intertwined. Thus some international legal instruments such as the International Labour Organization Convention (No. 169) concerning Indigenous and Tribal Peoples in Independent Countries of 1989 and the Convention on Biological Diversity, which was opened for signature at Rio, already specifically recognized and protected this relationship. Since Rio, indigenous peoples' special religious, cultural, indeed existential links with lands traditionally owned, occupied or used have been further clarified and given enhanced protection in a series of landmark decisions by human rights tribunals as well as in the United Nations Declaration on the Rights of Indigenous Peoples (General Assembly resolution 61/295).

j. Women in Development

The Rio Declaration was the very first international instrument to explicitly recognize that the empowerment of women and, specifically, their ability to effectively participate in their countries' economic and social processes, is an essential condition for sustainable development. Principle 20 of the Rio Declaration calls attention to women's "vital role in environmental management and development" and the consequent need for "their full participation." It recognizes the fact that women's livelihood, in particular in developing countries, often will be especially sensitive to environmental degradation. Unsurprisingly, this "women in development" perspective has been strongly endorsed in other international legal instruments, such as the preambles of the Convention on Biological Diversity or the Desertification Convention, and in resolutions of various international conferences. In short, as a United Nations Development Programme website puts it, gender equality and women's empowerment represent not only fundamental human rights issues, but "a pathway to achieving the Millennium Development Goals and sustainable development." However, as the calls for at Rio+20 seem to underline, much work appears still to be necessary before the Principle 20 objectives will truly be met.

k. Environmental Liability and Compensation

Finally, both the Stockholm and the Rio Declarations call for the further development of the law bearing on environmental liability and compensation. Whereas Stockholm Principle 22 refers to international law only, the corresponding Rio Principle 13 refers to both national and international law. Notwithstanding these clear mandates, States have tended to shy away from addressing the matter head-on or comprehensively, preferring instead to establish so-called private law regimes which focus on private actors' liability, while mostly excluding consideration of States' accountability. Recent developments, however, when taken together, can provide a basic frame of reference for issues related to environmental liability and compensation, be that at national or international level. These developments include, in particular, the work of the International Law Commission, especially its draft Principles on Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities; and the 2010 UNEP Guidelines for the Development of Domestic Legislation on Liability, Response Action and Compensation for Damage Caused by Activities Dangerous to the Environment. In this vein, therefore, it might be argued that today the expectations of legislative progress generated by the Stockholm and Rio Declarations have finally come to be realized, at least in large part.

THE EARTH SUMMIT AND AGENDA 21

Introduction

The United Nations Conference on Environment and Development (UNCED), which took place in Rio de Janeiro in June 1992, was a milestone event bringing together Heads of State and Chiefs of Government than any other meeting in the history of international relations, along with senior diplomats and government officials from around the globe, delegates from United Nations agencies, officials of international organizations, and many thousands of nongovernmental organization (NGO) representatives and journalists. UNCED made it plain that we can no longer think of environment and economic and social development as isolated fields. In addition to major international treaties and agreements concluded at the Earth Summit on issues of global climate change, biological diversity, deforestation, and desertification, the Declaration of Rio contains fundamental principles on which nations can base their future decisions and policies, considering the environmental implications of socio-economic development.

Agenda 21 was a special product of the Earth Summit. It is a vast work program for the 21st century, approved by consensus among the world leaders in Rio, representing over 98% of the world's population. This historic document is 700 pages long and embraces all areas of sustainable development. A comprehensive blueprint for a global partnership, Agenda 21 strives to reconcile the twin requirements of a high quality environment and a healthy economy for all people of the world, while identifying key areas of responsibility as well as offering preliminary cost estimates for success. The framing of Agenda 21 began well over a decade ago. By resolution 38/161 in December 1983, the UN General Assembly convened the World Commission on Environment and Development (WCED), chaired by Ms. Gro Harlem Brundtland, Prime Minister of Norway.

The 22 distinguished members of the WCED worked for three years, conducting a series of public hearings throughout the world, reviewing specially commissioned research and reports, and carrying on extensive international dialogue, to produce their unanimous report, Our Common Future, which was presented to the UN General Assembly in October 1987 and disseminated world-wide. The report placed the concept of sustainable development as an urgent imperative on the global agenda, and led directly to the decision by the United Nations to convene the 1992 Earth Summit. Agenda 21 reflects not only the testimony and counsel of the numerous technical and scientific advisers mobilized by the UNCED Secretariat under the leadership of Maurice F. Strong, but painstaking negotiation by the delegates of 172 sovereign nations. The Preparatory Committee, or PrepCom, held four month-long meetings from August 1990 through the spring of 1992. For deliberation at the Earth Summit, the 40 chapters of Agenda 21 were submitted in four sections to the corresponding four major committees of the delegates. Although Agenda 21 is a global consensus document, negotiation at Rio did not settle all disputes to the satisfaction of each participant...and not necessarily in the best interests of all, seen from the broadest perspective. It is, however, a unique step forward on the road toward sustainability, and offers a bold plan to mobilize local, national, and global action. Overview of Agenda 21 SECTION ONE: SOCIAL AND ECONOMIC DIMENSIONS The preamble and the following eight chapters consider the challenges that the adaptation of human behaviour to sustainable development pose to prevailing social and economic structures and institutions.

1. **PREAMBLE**

The preamble concludes, "Agenda 21 is a dynamic program. It will be carried out over time by the various actors according to the different situations, capacities and priorities of countries and regions...The process marks the beginning of a new global partnership..."

2. **ACCELERATING SUSTAINABLE DEVELOPMENT** Calls for a global partnership to provide a dynamic and growing world economy based on an "...open, equitable, secure, non-discriminatory, and predictable multilateral trading system," in which commodity exports of the developing countries can find markets at fair prices free of tariff and nontariff barriers. Cost: $8.8 billion

3**. COMBATING POVERTY**

Suggests that factors creating policies of development, resource management, and poverty be integrated. This objective is to be sought by improving access of the poor to education and health care, to safe water and sanitation, and to resources, especially land; by restoration of degraded resources; by empowerment of the disadvantaged, especially women, youth, and indigenous peoples; by ensuring that "women and men have the same right and the means to decide freely and responsibly on the number of spacing of their children." Cost of implementation: $30 billion 3.

4. **CHANGING CONSUMPTION PATTERNS**

"One of the most serious problems now facing the planet is that associated with historical patterns of unsustainable consumption, and production, particularly in the industrialized countries." Social research and policy should bring forward new concepts of status and lifestyles which are "less dependent on the Earth's finite resources and more in harmony with its carrying capacity." Greater efficiency in the use of energy and resources--for example, reducing wasteful packaging of products-- must be sought by new technology and new social values. Cost of implementation: The recommended measures are unlikely to require significant new financial resources.

5. **P OPULATION AND SUSTAINABILITY**

Urges governments to develop and implement population policies integral with their economic development programs. Health services should "include women-centered, women-managed, safe and effective reproductive health care and affordable, accessible services, as appropriate, for the responsible planning of family size..." Health services are to emphasize reduction of infant death rates which converge with low birth rates to stabilize world population at a sustainable number at the end of the century. Cost of implementation: $7 billion

6. PROTECTING AND PROMOTING HUMAN HEALTH

Calls for meeting basic health needs of all populations; provide necessary specialized environmental health services; co-ordinate involvement of citizens, and the health sector, in solutions to health problems. Health service coverage should be achieved for population groups in greatest need, particularly those living in rural areas. The preventative measures urged include reckoning with urban health hazards and risks from environmental pollution. Cost of implementation: $273 billion

7. SUSTAINABLE HUMAN SETTLEMENTS

Addresses the full range of issues facing urban-rural settlements, including: access to land, credit, and low-cost building materials by homeless poor and unemployed; upgrading of slums to ease the deficit in urban shelter; access to basic services of clean water, sanitation, and waste collection; use of appropriate construction materials, designs, and technologies; increased use of high-occupancy public transportation and bicycle and foot paths; reduction of long-distance commuting; support for the informal economic sector; development of urban renewal projects in partnership with non-governmental organizations; improved rural living conditions and land-use planning to prevent urban sprawl onto agricultural land and fragile regions. Cost of implementation: $218 billion 4.

8. MAKING DECISIONS FOR SUSTAINABLE DEVELOPMENT

Calls on governments to create sustainable development strategies to integrate social and environmental policies in all ministries and at all levels, including fiscal measures and the budget. Encourages nations and corporate enterprises to integrate environmental protection, degradation, and restoration costs in decision-making at the outset, and to mount without delay the research necessary to reckon such costs, to develop protocols bringing these considerations into procedures at all levels of decision-making. Cost of implementation: $63 million SECTION TWO: CONSERVATION AND MANAGEMENT OF RESOURCES The environment itself is the subject of chapters 9 through 22, dealing with the conservation and management of resources for development.

9. PROTECTING THE ATMOSPHERE

Urges constraint and efficiency in energy production and consumption, development of renewable energy sources; and promotion of mass transit technology and access thereto for developing countries. Conservation and expansion of "all sinks for greenhouse gases" is extolled, and transboundary pollution recognized as "subject to international controls." Governments need to develop more precise ways of predicting levels of atmospheric pollutants; modernize existing power systems to gain energy efficiency; and increase energy efficiency education and labelling programs. Cost of implementation: $21 billion

10. MANAGING LAND SUSTAINABLY

Calls on governments to develop policies that take into account the land-resource base, population changes, and the interests of local people; improve and enforce laws and regulations to support the sustainable use of land, and restrict the transfer of productive arable land to other uses; use techniques such as landscape ecological planning that focus on an ecosystem or a watershed, and encourage sustainable livelihoods; include appropriate traditional and indigenous land-use practices, such as pastoralism, traditional land reserves, and terraced agriculture in land management; encourage the active participation in decision-making of those affected groups that have often been excluded, such as women, youth, indigenous people, and other local communities; test ways of putting the value of land and ecosystems into national reports on economic performance; ensure that institutions that deal with land and natural resources integrate environmental, social, and economic issues into planning. Cost of implementation: $50 million 5.

11. COMBATING DEFORESTATION

Calls for concerted international research and conservation efforts to control harvesting of forests and "uncontrolled degradation and conversion to other types of land use," to develop the values of standing forests under sustained cultivation by indigenous technologies and agroforestry, and to expand the shrunken world-forest cover. Governments, along with business, nongovernmental and other groups should: plant more forests to reduce pressure on primary and old-growth forests; breed trees that are more productive and resistant to stress; protect forests and reduce pollutants that affect them, including air pollution that flows across borders; limit and aim to halt destructive shifting cultivation by addressing the underlying social and ecological causes; use environmentally sound, more efficient and less polluting methods of harvesting; minimize wood waste; promote small-scale enterprises; develop urban forestry for the greening of all places where people live; and encourage low-impact forest use and sustainable management of areas adjacent to forests. Cost of implementation: $31.25 billion

12. COMBATING DESERTIFICATION AND DROUGHT

Calls for intensive study of the process in its relation to world climate change to improve forecasting, study of natural vegetation succession to support large-scale revegetation and afforestation, checking and reversal of erosion, and like small-and grand-scale measures. For inhabitants whose perilously adapted livelihoods are threatened or erased, resettlement and adaptation to new life ways must be assisted. Governments must: adopt national sustainable land use plans and sustainable management of water resources; accelerate planting programs; and help to reduce the demand for fuelwood through energy efficiency and alternative energy programs. Cost of implementation: $8.6 billion.

13. S USTAINABLE MOUNTAIN DEVELOPMENT

Calls for study, protection, and restoration of these fragile ecosystems and assistance to populations in regions suffering degradation. Governments should: promote erosion-control measures that are low-cost, simple, and easily used; offer people incentives to conserve resources and use environment-friendly technologies; produce information on alternative livelihoods; create protected areas to save wild genetic material; identify hazardous areas that are most vulnerable to erosion floods, landslides, earthquakes, snow avalanches, and other natural hazards and develop early-warning systems and disaster-response teams; identify mountain areas threatened by air pollution from neighbouring industrial and urban areas; and create centres of information on mountain ecosystems. Cost of implementation: $13 billion.

14. S **USTAINABLE AGRICULTURE AND RURAL DEVELOPMENT**

Rising population food needs must be met through: increased productivity and co-operation involving rural people, national governments, the private sector, and the international community; wider access to techniques for reducing food spoilage, loss to pests, and for conserving soil and water resources; ecosystem planning; access of private ownership and fair market prices; advice and training in modern and indigenous conservation techniques including conservation tillage, integrated pest management, crop rotation, use of plant nutrients, agroforestry, terracing and mixed cropping; and better use and equitable distribution of information on plant and animal genetic resources. Cost of implementation: $30.8 billion

15. **CONSERVATION OF BIOLOGICAL DIVERSITY**

Recognizing the need to conserve and maintain genes, species, and ecosystems, urges nations, with the co-operation of the United Nations, nongovernmental organizations, the private sector, and financial institutions, to: conduct national assessments on the state of biodiversity; develop national strategies to conserve and sustain biological diversity and make these part of overall national development strategies; conduct long-term research into importance of biodiversity for ecosystems that produce goods and environmental benefits; protect natural habitats; encourage traditional methods of agriculture, agroforestry, forestry, range and wildlife management which use, maintain, or increase biodiversity. Cost of implementation: $3 billion.

16. MANAGEMENT OF BIOTECHNOLOGY

Calls for the transfer of biotechnology to the developing countries and the creation of the infrastructure of human capacity and institutions to put it to work there. Highlights need for internationally agreed principles on risk assessment and management of all aspects of biotechnology, to: improve productivity and the nutritional quality and shelf-life of food and animal feed products; develop vaccines and techniques for preventing the spread of diseases and toxins; increase crop resistance to diseases and pests, so that there will be less need for chemical pesticides; develop safe and effective methods for the biological control of disease-transmitting insects, especially those resistant to pesticides; contribute to soil fertility; treat sewage, organic chemical wastes, and oil spills more cheaply and effectively than conventional methods; and tap mineral resources in ways that cause less environmental damage. Cost of implementation: $20 billion.

17. PROTECTING AND MANAGING THE OCEANS

Sets out goals and programs under which nations may conserve "their" oceanic resources for their own and the benefit of the nations that share oceans with them, and international programs that may protect the residual commons in the interests even of land-locked nations, such as: anticipate 7. and prevent further degradation of the marine environment and reduce the risk of long-term or irreversible effects on the oceans; ensure prior assessment of activities that may have significant adverse impact on the seas; make marine environmental protection part of general environmental, social, and economic development policies; apply the "polluter pays" principle, and use economic incentives to reduce polluting of the seas; improve the living standards of coast-dwellers; reduce or eliminate discharges of synthetic chemicals that threaten to accumulate to dangerous levels in marine life; control and reduce toxic-waste discharges; stricter international regulations to reduce the risk of accidents and pollution from cargo ships; develop land-use practices that reduce run-off of soil and wastes to rivers, and thus to the seas; stop ocean dumping and the incineration of hazardous wastes at sea. Cost of implementation: $13 billion.

18. PROTECTING AND MANAGING FRESH WATER

Sets out measures, from development of long-range weather and climate forecasting to cleanup of the most obvious sources of pollution, to secure the supply of fresh water for the next doubling of the human population. Focus is on developing low-cost but adequate services that can be installed and maintained at the community level to achieve universal water supply by 2025. The interim goals set for 2000 include: to provide all urban residents with at least 40 liters of safe drinking water per person per day; provide 75% of urban dwellers with sanitation; establish standards for the discharge of municipal and industrial wastes; have three-quarters of solid urban waste collected and recycled, or disposed of in an environmentally safe way; ensure that rural people everywhere have access to safe water and sanitation for healthy lives, while maintaining essential local environments; control water-associated diseases. Cost of implementation $54.7 billion.

19. SAFER USE OF TOXIC CHEMICALS

Seeks objectives such as: full evaluation of 500 chemicals before the year 2000; control of chemical hazards through pollution prevention, emission inventories, product labelling; use limitations, procedures for safe handling and exposure regulations; phase-out or banning of high-risk chemicals; consideration of policies based on the principle of producer liability; reduced risk by using less-toxic or non-chemical technologies; review of pesticides whose acceptance was based on criteria now recognized as insufficient or outdated; efforts to replace chemicals with other pest-control methods such as biological control; provision to the public of information on chemical hazards in the languages of those who use the materials; development of a chemical-hazard labelling system using easily understandable symbols; control of the export of banned or restricted chemicals and provision of information on any exports to the importing countries. Cost of implementation: $600 million.

20. MANAGING HAZARDOUS WASTES

Seeks international support in restraint of the trade and for containing the hazardous cargoes in safe sinks. Governments should: require and assist in the innovation by industry of cleaner production methods and of preventive and recycling technologies; encourage the phasing out of processes that produce high risks because of hazardous waste management; hold producers responsible for the environmentally unsound disposal of the hazardous wastes they generate; establish public information programs and ensure that training programs provided for industry and government workers on hazardous-waste issues, especially use minimization; build treatment centres for hazardous wastes, either at the national or regional level; ensure that the military conforms to national environmental norms for hazardous-waste treatment and disposal; ban the export of hazardous wastes to countries that are not equipped to deal with those wastes. Industry should: treat, recycle, reuse, and dispose of wastes at or close to the site where they are created. Cost of implementation: $18.5 billion

21. MANAGING SOLID WASTES AND SEWAGE

Governments should urge waste minimization and increased reuse/recycling as strategies toward sound waste treatment and disposal; encourage "life-cycle" management of the flow of material into and out of manufacturing and use; provide incentives to recycling; fund pilot programs, such as small-scale and cottage-based recycling industries, compost production, irrigation using treated waste water, and the recovery of energy from wastes; establish guidelines for the safe reuse of waste and encourage markets for recycled and reused products. Cost of implementation: $23.3 billion

21. MANAGING RADIOACTIVE WASTES

Calls for increasingly stringent measures to encourage countries to co-operate with international organizations to: promote ways of minimizing and limiting the creation of radioactive wastes; provide for the sage storage, processing, conditioning, transportation, and disposal of such wastes; provide developing countries with technical assistance to help them deal with wastes, or make it easier for such countries to return used radioactive material to suppliers; promote the proper planning of safe and environmentally sound ways of managing radioactive wastes, possibly including assessment of the environmental impact; strengthen efforts to implement the Code of Practice on the Transboundary Movements of Radioactive Wastes; encourage work to finish studies on whether the current voluntary moratorium on disposal of low-level radioactive wastes at sea should be replaced by a ban; not promote or allow storage or disposal of radioactive wastes near seacoasts or open seas, unless it is clear that this does not create an unacceptable risk to people and the marine environment; not export radioactive wastes to countries that prohibit the import of such waste. Cost of implementation: $8 million.

SECTION THREE: STRENGTHENING THE ROLE OF MAJOR GROUPS

The issues of how people are to be mobilized and empowered for their various roles in sustainable development are addressed in chapters 23 through 32.

23. **PREAMBLE**

"Critical to the effective implementation of the objectives, policies, and mechanisms agreed to by Governments in all program areas of Agenda 21 will be the commitment and involvement of all social groups..."

24. WOMEN IN SUSTAINABLE DEVELOPMENT

Urges governments to face the status question; give girls equal access to education; reduce the workloads of girls and women; make health-care systems responsive to female needs; open employment and careers to women; and bring women into full participation in social, cultural, and public life. Governments should: ensure a role for women in national and international ecosystem management and control of environmental degradation; ensure women's access to property rights, as well as agricultural inputs and implements; take all necessary measures to eliminate violence against women, and work to eliminate persistent negative images, stereotypes, and attitudes, and prejudices against women; develop consumer awareness among women to reduce or eliminate unsustainable consumption; and begin to count the value of unpaid work. Cost of implementation: $40 million.

25. CHILDREN AND YOUTH IN SUSTAINABLE DEVELOPMENT

Calls on governments, by the year 2000, to ensure that 50% of their youth, gender balanced, have access to secondary education or vocational training; teach students about the environment and sustainable development through their schooling; consult with and let youth participate in decisions that affect the environment; enable youth to be represented at international meetings, and participate in decision-making at the United Nations; combat human rights abuses against youth and see that their children are healthy, adequately fed, educated, and protected from pollution and toxic substances; and develop strategies that deal with the entitlement of young people to natural resources. Cost of implementation: $1.5 million.

26. STRENGTHENING THE ROLE OF INDIGENOUS PEOPLES

Urges governments to enrol indigenous peoples in full global partnership, beginning with measures to protect their rights and conserve their patrimony; recognize that indigenous lands need to be protected from environmentally unsound activities, and from activities the people consider to be 10. socially and culturally inappropriate; develop a national dispute resolution procedure to deal with settlement and land-use concerns; incorporate their rights and responsibilities into national legislation; recognize and apply elsewhere indigenous values, traditional knowledge and resource management practices; and provide indigenous people with suitable technologies to increase the efficiency of their resource management. Cost of implementation: $3 million.

27. PARTNERSHIPS WITH NONGOVERNMENTAL GROUPS [CIVIC GROUPS]

Calls on governments and the United Nations system to: invite nongovernmental groups to be involved in making policies and decisions on sustainable development; make NGOs a part of the review process and evaluation of implementing Agenda 21; provide NGOs with timely access to information; encourage partnerships between NGOs and local authorities; review financial and administrative support for NGOs; utilize NGO expertise and information; and create laws enabling NGOs the right to take legal action to protect the public interest. Cost of implementation: no estimate.

28. LOCAL AUTHORITIES

Calls on local authorities, by 1996, to undertake to promote a consensus in their local populations on "a local Agenda 21;" and, at all times, to invite women and youth into full participation in the decision-making, planning, and implementation process; to consult citizens and community, business, and industrial organizations to gather information and build a consensus on sustainable development strategies. This consensus would help them reshape local programs, policies, laws, and regulations to achieve desired objectives. The process of consultation would increase people's awareness of sustainable development issues. Cost of implementation: $1 million.

29. WORKERS AND TRADE UNIONS

Challenges governments, businesses, and industries to work toward the goal of full employment, which contributes to sustainable livelihoods in safe, clean, and healthy environments, at work and beyond, by fostering the active and informed participation of workers and trade unions in shaping and implementing environment and development strategies at both the national and international levels; increase worker education and training, both in occupational health and safety and in skills for sustainable livelihoods; and promote workers' rights to freedom of association and the right to organize. Unions and employees should design joint environmental policies, and set priorities to improve the working environment and the overall environmental performance of business and develop more collective agreements aimed at achieving sustainability. Cost of implementation: $300 million.

30. BUSINESS AND INDUSTRY

Calls on governments to: use economic incentives, laws, standards, and more streamlined administration to promote sustainably managed enterprises with cleaner production; encourage the creation of venture-capital funds; and co-operate with business, industry, academia, and international organizations to support training in the environmental aspects of enterprise management. Business and industry should: develop policies that result in operations and products that have lower environmental impacts; ensure responsible and ethical management of products and processes from the point of view of health, safety, and the environment; make environmentally sound technologies available to affiliates in developing countries without prohibitive charges; encourage overseas affiliates to modify procedures in order to reflect local ecological conditions and share information with governments; create partnerships to help people in smaller companies learn business skills; establish national councils for sustainable development, both in the formal business community and in the informal sector, which includes small-scale businesses, such as artisans; increase research and development of environmentally sound technologies and environmental management systems; report annually on their environmental records; and adopt environmental and sustainable development codes of conduct. Cost of implementation: no estimate.

31. SCIENTISTS AND TECHNOLOGISTS

Indicates that governments should: decide how national scientific and technological programs could help make development more sustainable; provide for full and open sharing of information among scientists and decision-makers; fashion national reports that are understandable and relevant to local sustainable development needs; form national advisory groups to help scientists and society develop common values on environmental and developmental ethics; and put environment and development ethics into education and research priorities. Scientists and technologies have special responsibilities to: search for knowledge, and to help protect the biosphere; increase and strengthen dialogue with the public; and develop codes of practice and guidelines that reconcile human needs and environmental protection. Cost of implementation: $20 million.

32. STRENGTHENING THE ROLE OF FARMERS

To develop sustainable farming strategies, calls on governments to collaborate with national and international research centres and nongovernmental organizations to: develop environmentally sound farming practices and technologies that improve crop yields, maintain land quality, recycle nutrients, conserve water and energy, and control pests and weeds; help farmers share expertise in conserving land, water, and forest resources, making the most efficient use of chemicals and reducing or re-using farm wastes; encourage self-sufficiency in low-input and low-energy technologies, including indigenous practices; support research on equipment that makes optimal use of human labour and animal power; delegate more power and responsibility to those who work the 12. land; give people more incentive to care for the land by seeing that men and women can get land tenure, access to credit, technology, farm supplies, and training. Researchers need to develop environment-friendly farming techniques and colleges need to bring ecology into agricultural training. Cost of implementation: no estimate. SECTION FOUR: MEANS OF IMPLEMENTATION Chapters 33 through 40 deal with the ways and means of implementing Agenda 21.

33. FINANCING SUSTAINABLE DEVELOPMENT

At UNCED, countries committed to the consensus of a global partnership, holding that the eradication of poverty "is essential to meeting national and global sustainability objectives;" that "the cost of inaction could outweigh the financial costs of implementing Agenda 21;" that "the huge sustainable development programs of Agenda 21 will require the provision to developing countries of substantial new and additional financial resources;" and that "the initial phase will be accelerated by substantial early commitments of concessional funding." Further, the developed countries "reaffirmed their commitments to reach the accepted United Nations target of 0.7% of GNP for concessional funding... as soon as possible." Cost of implementation: $561.5 billion per year total for all programs, including $141.9 billion in concessional financing.

34. TECHNOLOGY TRANSFER

Economic assistance would move from the developed to the developing counties principally in the form of technology. Developing countries would be assisted in gaining access to technology and know-how in the public domain and to that protected by intellectual property rights as well, "taking into account developments in the process of negotiating an international code of conduct on the transfer of technology" proceeding under the United Nations Agreement on Tariffs and Trade. To enhance access of developing countries to environmentally sound technology, a collaborative network of laboratories is to be established. Cost of implementation: $500 million

35. SCIENCE FOR SUSTAINABLE DEVELOPMENT

Sustainable development requires expansion of the on-going international collaborative enterprises in the study of the geochemical cycles of the biosphere and the establishment of strong national scientific enterprises in the developing countries. The sciences link fundamental understanding of the Earth system to development of strategies that build upon its continued healthy functioning. "In the face of threats of irreversible environmental damage, lack of full scientific understanding should not be an excuse for postponing actions which are justified in their own right." 13. Countries need to develop tools for sustainable development, such as: quality-of-life indicators covering health, education, social welfare, and the state of environment, and the economy; economic incentives that will encourage better resource management; and ways of measuring the environmental soundness of new technologies. They should use information on the links between the state of ecosystems and human health when weighing the costs and benefits of different development policies, and conduct scientific studies to help map our national and regional pathways to sustainable development. When sustainable development plans are being make, the public should be involved in setting long-term goals for society. Cost of implementation: $3 billion.

36. EDUCATION, TRAINING, AND PUBLIC AWARENESS

Because sustainable development must ultimately enlist everyone, access to education must be hastened for all children; adult illiteracy must be reduced to half of its 1990 level, and the curriculum must incorporate environmental and developmental learning. Nations should seek to: introduce environment and development concepts, including those related to population growth, into all educational programs, with analyses of the causes of the major issues. They should emphasize training decision-makers; involve schoolchildren in local and regional studies on environmental health, including safe drinking water, sanitation, food, and the environmental and economic impacts of resource use; set up training programs for school and university graduates to help them achieve sustainable livelihoods; encourage all sectors of society to train people in environmental management; provide locally trained and recruited environmental technicians to give local communities services they require, starting with primary environmental care; work with the media, theatre groups, entertainment, and advertising industries to promote a more active public debate on the environment; and bring indigenous peoples' experience and understanding of sustainable development into education and training. Cost of implementation: $14.6 billion.

37. CREATING CAPACITY FOR SUSTAINABLE DEVELOPMENT

Developing countries need more technical co-operation and assistance in setting priorities so that they can deal with new long-term challenges, rather than concentrating only on immediate problems. For example, people in government and business need to learn how to evaluate the environmental impact of all development projects, starting from the time the projects are conceived. Assistance in the form of skills, knowledge, and technical know-how can come from the United Nations, national governments, municipalities, nongovernmental organizations, universities, research centres, and business and other private organizations. The United Nations Development Program has been given responsibility for mobilizing international funding and co-ordination programs for capacity building. Cost of implementation: $650 million.

38. ORGANIZING FOR SUSTAINABLE DEVELOPMENT

To the existing UN system, the General Assembly as the supreme deliberative and policymaking body, the Economic and Social Council as the appropriate overseer of system-wide coordination reporting to the General Assembly, the Secretary General as chief executive, and the technical agencies seeing to their special functions, Agenda 21 proposes to add a Commission on Sustainable Development to monitor implementation of Agenda 21, reporting to the General Assembly through ECOSOC. The Conference also recommended that the UN Secretary-General appoint a high-level board of environment and development experts to advise on other structural change required in the UN system. The United Nations Environment Program will need to develop and promote natural resource accounting and environmental economics, develop international environmental law, and advise governments on how to integrate environmental considerations into their development policies and programs. Cost of implementation: no estimate.

39. INTERNATIONAL LAW

The major goals in international law on sustainable development should include: the development of universally negotiated agreements that create effective international standards for environmental protection, taking account of the different situations and abilities of various countries; an international review of the feasibility of establishing general rights and obligations of nations as in the field of sustainable development; and measures to avoid or settle international disputes in the field of sustainable development. These measures can range from notification and talks on issues that might lead to disputes, to the use of the International Court of Justice. Cost of implementation: no estimate.

40. INFORMATION FOR DECISION -MAKING

Calls on governments to ensure that local communities and resource users get the information and skills needed to manage their environment and resources sustainably, including application of traditional and indigenous knowledge; more information about the status of urban air, fresh water, land resources, desertification, soil degradation, biodiversity, the high seas, and the upper atmosphere; more information about population, urbanization, poverty, health, and rights of access to resources. Information is also needed about the relationships of groups, including women, indigenous peoples, youth, children and the disabled with environment issues. Current national accounting reckons environmental costs as "externalities." Internalization of such costs, the amortization of non-renewable resources, and the development of indicators of sustainability all require not only new data but new thinking.

Earth Summit: 2012

- **Introduction**

The overall aim of this ESDN Quarterly Report is to present and review the preparation phase and outcomes of the United Nations Conference on Sustainable Development (UNCSD) that took place in Rio de Janeiro from 20-22 June 2012. The conference was hosted 20 years after the crucial United Nations Conference on Environment and Development (UNCED) held in Rio in 1992, also known as the 1992 Earth Summit. For this reason, the report will refer to it as the Rio+20 Conference, as it is also commonly addressed by the UN and in the media. Before we go into the details of sketching the road to Rio+20 and the analysis of the Rio+20 process and outcome document, we elaborate in the next chapter on global sustainable development (SD) governance as the conceptual framework for the Rio+20 Conference. 2 Global Sustainable Development Governance 2.1 Definition of global SD governance Finding a useful and commonly accepted definition for "global SD governance" is not an easy task. Controversies could certainly emerge either regarding the definition of 'sustainable development' or the concept of 'governance' (Waas et al., 2011; Steurer, 2009; George, 2007; Hanson, 2007; Baker, 2006; Sneddon et al., 2006; Redclift, 2005; Robinson, 2004). We nevertheless find it useful to present some basic definitions of both: Governance usually "refers to the managing, steering and guiding of public affairs by governing procedures and institutions in a democratic manner, especially in relation to public policy decision-making" (Baker, 2009; Jordan, 2008; Lafferty, 2004).

Furthermore, 'governance for SD' can be defined as "processes of socio-political governance oriented towards the attainment of sustainable development. It encompasses public debate, political decision-making, policy formation and implementation, and complex interactions among public authorities, private business and civil society – in so far as these relate to steering societal development along more sustainable lines" (Meadowcroft, 2007, p.299). Global SD governance is, therefore, understood as organised action of institutions, organisations, communities and individuals taken in order to achieve SD objectives concerning issues that have international facets and global interests (Hanson, 2008). In this context, national governments and subnational levels, intergovernmental and transnational bodies, businesses and industries, nongovernmental and civil-society organisations, communities and individuals can have a stake in the promotion of SD.

Over the last 40 years, a crucial role in global SD governance has been played by the United Nations (UN), especially through a series of international summits and conferences: these forums haveundoubtedly shaped what today is referred as global SD governance. Notwithstanding a number of backdrops (see for instance, Seyfang and Jordan, 2002), the UN system and UN conferences have had a number of positive functions, such as those identified by Haas (2002) and Baker (2006): agenda setting; popularizing issues; raising consciousness; generating new information and new challenges for government; providing general alerts and warning of new threats; galvanizing administrative reform; adopting new norms and doctrinal consensus; promoting mass involvement. Others, like Death (2011) or Swyngedouw (2010), are more sceptical about global SD and environmental summits and conferences. While acknowledging that these events have become firmly established as "landmark moments" of SD and environmental governance, Death (2011, p.2) argues that they reinforce "dominant hierarchical, state-centric, elitist and rationalist models of politics" and mainly have "symbolic, performative and theatrical roles". This also implies, Death further argues, that symbolic politics and public relations, which are key to heads-of-state and other politicians, dominate rather than addressing the complex nature of SD, offer short-term solutions, or establishing political visions.

In other words, as Swyngedouw (2010, p. 223) put it, political elites at such global mega-summit stages tend to "undertake action such that nothing really has to change, so that life can basically go on as before" and these events fail to "create real possibilities for constructing different socio-environmental futures" (ibid., p. 228). However, "there is little doubt that the UN environment summits and conferences have contributed to ... [establishing] the agenda of global environmental politics around the aim of promoting sustainable development" (Baker, 2006, p.73). Moreover, as Death (2011, p.2) mentions, global summits play a crucial role in showing global audiences that political elites are serious about issues like sustainable development and climate change and that they still offer, maybe paradoxically, a platform for non-state actors to engage in and demonstrate for/against respective conference topics.

Bodies of Global SD Governance

In the UN system, a high number of international institutions, organisations and bodies have been created over the last 40 years to foster global SD governance, and this is also why "the UN system for the promotion of SD penetrates into all areas of international governance" In Rogers et. al. (2008), an impressive list of more than 20 institutions participating in the global SD governance are depicted (p.348): 1)United Nations (UN) 2)United Nations Environment Programme (UNEP) 3)United Nations Development Programme (UNDP) 4) Food and Agriculture Organization (FAO) 5)International Labor Organization (ILO) 6)United Nations Education, Scientific and Cultural Organization (UNESCO) 7)United Nations Conference on Trade and Development (UNCTAD) 8)World Health Organization (WHO) 9)World Meteorological Organization (WMO) 10) International Atomic Energy Agency (IAEA) 11) World Bank (WB) 12) World Trade Organization (WTO) 13) United Nations Department of Economic and Social Affairs (UN/DESA) 14) United Nations Children's Fund (UNICEF) 15) World Conservation Union (IUCN) 16) World Wildlife Fund for Nature (WWF) 17) United Nations Commission for Sustainable Development (UNCSD) 18) Interagency Committee on Sustainable Development (IACSD) 19) Regional development banks (RDBs), 20) International Fund for Agricultural Development (IFAD), 21) Bilateral agencies, 22) NGOs, 23) Private foundations. Notwithstanding the importance of all UN bodies, in 2011, the Stakeholder Forum and the Commonwealth Secretariat pointed out the following seven as the most relevant UN bodies responsible for implementing SD1 .

The 7 Most Relevant UN Bodies for SD Implementation

UN General Assembly:

This is the key operational body of the United Nations, with functions of deliberation, policy-making and representation. Through its adoption of SD-related resolutions (i.e. Rio Declaration) which "reaffirm the connection in the fields of economic, social and environmental development (…) sustainable development becomes a central element in the UN framework. Subsequently, the General Assembly deals with sustainable development in the process of standard setting, draft laws and regulation, as well in the implementation of measures adopted. Finally, it liaises with all other UN bodies in order to achieve improved coordination of UN activities on sustainable development-related issues" (p.27);

Second Committee or Economic and Financial Committee (ECOFIN):

Although mainly concerned with macroeconomic issues, the Second Committee is a committee within the United Nations that addresses issues in the areas of global finances and economics, including issues relating to international trade, financing for development, sustainable development and poverty.

Economic and Social Council (ECOSOC):

It consists of 54 UN member states, elected by the General Assembly. Its function is to restructure and revitalise UN activities in economic, social and related fields and manage sustainable development coordination within the UN system, integrating environmental and developmental issues within UN policies and programmes. ECOSOC is also in charge of undertaking studies and publishing reports on international issues of development, health, education, and sustainable development, among others, and making recommendations on such issues to the General Assembly, UN members and specialised agencies; eradication;

UN Commission on Sustainable Development (UNCSD):

It was established by the UN General Assembly in December 1992 to ensure effective follow-up of United Nations Conference on Environment and Development (UNCED), as a functioning commission of the UN Economic and Social Council (ECOSOC). The Secretariat functions of the CSD are performed by the Division for Sustainable Development (DSD) which has a broader remit to provide leadership and an authoritative source of expertise within the UN on sustainable development. In turn the DSD resides within the UN Department for Economic and Social Affairs (UNDESA), whose mission is to promote 'development for all', with sub-divisions focusing on particular elements of that vision. The CSD acts as a coordination organ for sustainable development issues at ECOSOC. Its main function is to monitor progress towards internationally agreed goals on sustainable development, and to enhance dialogue among and between governments, NGOs, UN agencies and other stakeholders, and to make recommendations to the General Assembly via ECOSOC;

Environmental Management Group (EMG):

Is an inter-agency coordinating body for environmental issues across the UN system with the objective to address inefficiencies and overlaps in the environmental governance system. The EMG coordinates and supports many UN consultative processes that aim to further understanding on environmental governance and to develop approaches that result in sound cooperation at an international level. The EMG also plays an important role in enhancing coherence and mainstreaming environmental considerations at a country level, through operational activities. There are clear and necessary responsibilities on national governments to promote a coherent national governance framework for multilateral environmental obligations. In supporting countries at the operational level, the EMG can also help to improve States' approaches to environmental governance

UN Environmental Programme (UNEP):

Founded as a result of the United Nations Conference on the Human Environment in Stockholm in June 1972, UNEP is responsible for the coordination of environmental activities across the UN system, assisting developing countries in implementing environmentally sound policies and practices. Seen as 'the voice for the environment within the United Nations System', UNEP is a cross-sectoral body which works in partnership with a range of actors, i.e. UN bodies, international organisations, non-governmental organisations, the private sector, etc. UNEP promotes environmental protection and the sustainable use of global natural resources by providing funding, education, facilitating multilateral discussion and pushing forward international environmental regimes. Furthermore, a large part of UNEP's recent activities focus on understanding, mitigating and adapting to climate change. UNEP established the Intergovernmental Panel on Climate Change (IPCC) together with the World Meteorological Organisation;

UN Development Programme (UNDP):

'An organization advocating for change and connecting countries to knowledge, experience and resources to help build a better life', the UNDP is the UN's global development network. A major focus of the UNDP is poverty alleviation and the achievement of the Millennium Development Goals, focusing on the following areas: i) Democratic Governance, ii) Poverty Reduction, iii) Crisis Prevention and Recovery, iv) Environment and Energy, v) HIV/AIDS. UNDP was identified in Agenda 21 as one of the core agencies for delivering sustainable development. The UNDP is a critical global institution for the implementation of sustainable development, partly because it focuses on issues that are highly relevant to all three pillars of sustainable development, but also because it has the capacity, budget, and global reach to significantly advance sustainable development objectives. Though UNDP primarily represents the 'social' pillar of sustainable development, it also focuses heavily on the environmental pillar through its Environment and Energy programme, and can play a key role in integrating the three pillars of sustainable development at a global level.

Stakeholder Forum for a Sustainable Future

Stakeholder Forum is an international not-for-profit organisation working to advance sustainable development and promote democracy at a global level. Its aims are to enhance open, accountable and participatory international decision-making on sustainable development through enhancing the involvement of stakeholders in intergovernmental processes while it seeks to provide a bridge between those who have a stake in sustainable development, and the international forums where decisions are made in their name. Founded in 1987 as UNED UK – United Nations Environment and Development UK – it was renamed Stakeholder Forum for a Sustainable Future in 2000 to reflect the broad range of activities that the organization undertakes. Stakeholder Forum is the organisation leading on global stakeholder engagement and is also the leading organisation in developing and facilitating global multi-stakeholder processes on sustainable development.

Asian Development Bank (ADB)

Since its founding in 1966, ADB has been driven by an inspiration and dedication to improving people's lives in Asia and the Pacific. Whether through investment in infrastructure, health care services, financial and public administration systems, or helping nations prepare for the impact of climate change or better manage their natural resources, ADB is committed to helping developing member countries evolve into thriving, modern economies that are well integrated with each other and the world.

Northern Alliance for Sustainability (ANPED)

ANPED is an international not-for-profit organization representing a vast network of NGOs in the Northern hemisphere with a mission to pro-actively promote the agenda on environmental justice and systemic change for the Economy. We also empower Northern civil society through capacity development, exchanges and knowledge sharing while working in close cooperation with Southern civil society and other stakeholders, for the creation and protection of sustainable societies worldwide. ANPED is the UN officially appointed Organizing Partner (OP) for the Major Group of the NGOs and disseminate information about initiatives, which support preparations for the UN CSD Rio+20.

Pardee Center at the Boston University

The Frederick S. Pardee Center for the Study of the Longer-Range Future convenes symposia and conducts interdisciplinary, policy-relevant, and future-oriented research that contributes to long-term improvements in the human condition. The Center's focus is defined by its longerrange vision and it seeks to identify, anticipate, and enhance the long-term potential for human progress— with recognition of its complexity and uncertainties.

International Institute for Sustainable Development (IISD)

The International Institute for Sustainable Development (IISD) is a Canadian-based, public policy research institute that has a long history of conducting cutting-edge research into sustainable development. Today, the institute is a non-partisan, charitable organization specializing in policy research, analysis and information exchange. The institute champions global sustainable development through innovation, research and relationships that span the entire world. It is devoted to the ongoing communication of its findings as it engages decision-makers in business, government, non-government organizations and other sectors.

MacArthur Foundation

The John D. and Catherine T. MacArthur Foundation supports creative people and effective institutions committed to building a more just, verdant, and peaceful world. In addition to selecting the MacArthur Fellows, the Foundation works to defend human rights, advance global conservation and security, make cities better places, and understand how technology is affecting children and society. Through the support it provides, the Foundation fosters the development of knowledge, nurtures individual creativity, strengthens institutions, helps improve public policy, and provides information to the public, primarily through support for public interest media.

In addition, international declarations and principles have been established and are still shaping what today is referred as 'global SD governance'. While the most prominent definition of SD – a "development, that meets the needs of the present without compromising the ability of future generations to meet their own needs" (WCED, 1987) – is provided by the Brundtland report, it is the Rio Declaration (UNCED, 1992) that offers the most recognised list of principles for global SD governance. Among the 27 principles contained in the Rio Declaration, the following 6 are generally regarded as the most significant ones for SD governance:

Inter- and intra-generational equity (Principle 3)

"The right to development must be fulfilled so as to equitably meet developmental and environmental needs of present and future generations";

Environmental Policy Integration (Principle 4)

"In order to achieve sustainable development, environmental protection shall constitute an integral part of the development process and cannot be considered in isolation from it";

Common but differentiated responsibilities (Principle 7)

“States shall cooperate in a spirit of global partnership to conserve, protect and restore the health and integrity of the Earth's ecosystem. In view of the different contributions to global environmental degradation, States have common but differentiated responsibilities. The developed countries acknowledge the responsibility that they bear in the international pursuit to sustainable development in view of the pressures their societies place on the global environment and of the technologies and financial resources they command”;

- **Public participation (Principle 10)**

“Environmental issues are best handled with participation of all concerned citizens, at the relevant level. At the national level, each individual shall have appropriate access to information concerning the environment that is held by public authorities, including information on hazardous materials and activities in their communities, and the opportunity to participate in decisionmaking processes. States shall facilitate and encourage public awareness and participation by making information widely available. Effective access to judicial and administrative proceedings, including redress and remedy, shall be provided”;

The precautionary principle (Principle 15)

“In order to protect the environment, the precautionary approach shall be widely applied by States according to their capabilities. Where there are threats of serious or irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation”;

Polluter pays principle (Principle 16)

“National authorities should endeavour to promote the internalization of environmental costs and the use of economic instruments, taking into account the approach that the polluter should, in principle, bear the cost of pollution, with due regard to the public interest and without distorting international trade and investment”.

The Johannesburg Summit 2002

Despite the positive effects in the aftermath and the results of UNCED, governments have not been implementing the plans for action the UNCED and other conferences have produced. Thus the World Summit on Sustainable Development (WSSD) in Johannesburg in 2002 convened in order to reinvigorate the global commitment to sustainable development and review the progress. In this regard, the report assessing the implementation status of Agenda 21 identified some serious deficiencies: fragmented approach to SD; lack of progress in addressing unsustainable patterns of consumption and production; inadequate attention of core issues (WEHAB); coherence policies on finance, trade, investment, technology and SD; insufficient financial resources; and absence of a robust mechanism for technology transfer (Hens & Nath, 2005).

The Johannesburg Summit delivered three outcomes: a political declaration, the Johannesburg Plan of Implementation, and the establishment of numerous partnership initiatives (most importantly so-called Type II partnerships). The Johannesburg Declaration confirmed the commitments from Stockholm 1972 and Rio 1992 as well as of some of the Millennium Development Goals. Another important deliverable – the Johannesburg Plan of Implementation (JPOI) – can be regarded as a programme of action to guide government activities, negotiated and agreed between governments covering key commitments and targets in the areas of sustainable consumption and production, water and sanitation, and energy. In addition, the Johannesburg Summit 2002 produced so-called Type II Partnerships (i.e.

voluntary transnational multi-stakeholder agreements between government and non-state actors) allowing civil society to contribute to the implementation of sustainable development.

Although the Johannesburg Declaration refers to strategic approaches on how to deliver the Johannesburg Plan of Implementation, it had no specific mandate to contribute to the development of international environmental law, nor even to further elaborate general principles of non-binding nature to guide the conduct of states with respect to SD. (von Frantzius, 2007; Hens & Nath, 2005). Negotiations on the document were rather short due to time constraints and, especially the part on implementation with regard to finance and trade was heavily controversial with an agreement of about 11 and 15 %, respectively (Hens & Nath, 2005).

The **Johannesburg Plan of Implementation (JPOI)** – a guide for further implementation of Agenda 21 – comprises measures of implementation and specific measureable targets and associated time frames, which, however, are in most cases reiterated from the Millennium Development Goals (MDGs) agreed at the Millennium Summit in 2000 and other agreements (Bigg, 2003). Nevertheless, according to Hens & Nath (2005), by doing so the JPOI confirmed and lifted some of these targets to a higher level of international agreement in response to lack of progress to date. Type II partnerships are generally perceived as powerful tools and more democratic instruments for the implementation of Agenda 21, however, they are also seen as an ambivalent instrument, as they offer the 'possibility to deliver some results without really committing governments to hard action' (Heinrich Böll Stiftung, 2002).

Consequently, it remains unclear whether the formal recognition of **type II partnerships** would strengthen the principal inter-governmental commitments or marginalise them. Von Frantzius (2007) goes one step further by claiming that these partnerships potentially mask the failure of governments resulting in a 'privatisation of sustainable development', whereas Bäckstrand (2006) perceives these instruments as key innovations in a shift from purely top-down to bottom-up governance approaches. Overall, the discussions that took place during the conference shifted the attention of SD away from the environmental and more towards the social and economic development perspective. This shift was mainly driven by the developing countries' needs and particularly influenced by the Millennium Declaration and its associated goals partly reiterated into the conference's final deliverables. In this sense, Rajamani (2003) argues that this shift is reflected by the introduction of the principle of common but differentiated responsibilities and the focus on issues such as poverty, education, sanitation.

Conversely, Seyfang (2003) and von Frantzius (2007) experience the inclusion of poverty alleviation and eradication in the JPOI as an important step towards integrating social and economic aspects of sustainable development with environmental goals. Despite the conferences efforts to integrate civil society in the negotiations of the outcome (i.e. Multistakeholder dialogues), they remained rather secondary to traditional state-centric negotiations or public relations exercises (Death, 2011). The failure of multi-stakeholder involvement can be explained by the facts that they (1) came too late in the negotiation process for being substantially influential (Karlsson, 2012); (2) were characterized by a lack of participation from government delegations (Hiblin et al., 2002; IISD, 2002); (3) and had rather formal monologue style of discussion than being an interactive dialogue (Bäckstrand, 2006). In this sense, the multi-stakeholder dialogues remained a rather symbolic gesture of global democratic governance within the negotiation process and, according to Karlsson (2012), 'final negotiations were as usual dominated by diplomats defending their country's particular priorities and sensitivities linked to national sovereignty'. However, the institutionalization of new and innovative deliberative practices, according to Bäckstrand (2006), represented a shift toward enacting a model which reflects key features of the stakeholder model of democracy (i.e. deliberative democracy) and which Sneddon et al. (2006) believe 'is crucial to any discussion of SD policies and sustainability politics'.

Rio+20 – Objectives, Negotiations and Outcomes

The aim of this chapter is to present the preparation and outcomes of the Rio+20 Conference in detail and to complement studies already produced. An overview of the conference, its crucial objectives and themes will be offered, together with insights on the process. Most of the information is retrieved from the Rio+20 website and from other official sources, especially United Nations. Particular attention will be also devoted to the EU position. In addition, we offer a word cloud analysis of the final 'draft' Rio+20 text, "The Future We Want", that was presented to the heads-of-state (version of 19 June), with the final texts of the Rio declaration 1992 and Johannesburg Plan of Implementation 2002. 4.1 Rio+20: the conference start "We have enough papers; we have enough conferences.

What we need to do now is something really different: Rio+20 should be not another conference in normal sense; it should be a conference of action, a conference of implementation of what we have agreed twenty or ten years ago." Sha Zukang (Conference Secretary-General for Rio+20)3 Commonly known as the Rio+20 Conference, the United Nations Conference on Sustainable Development (UNCSD) took place in Rio de Janeiro between the 20-22 June 2012, twenty years after the UNCED (United Nations Conference on Environment and Development), which was also hosted in Rio in 1992. More than 150 heads-of-state and ministers kicked off the Rio+20 Conference on 20 June 2012. Important heads-of-state, like Barack Obama (USA), David Cameron (UK), and Angela Merkel (Germany) as well as dozens of other leaders "have snubbed the talks" (The Guardian, 20 June 2012, a4).

The most important heads-of-state were, as the New York Times put it, "preoccupied by domestic politics and the financial turmoil in Europe" (New York Times, 18 June5). The Guardian argued that "the absence of so many key figures has dismayed the architects of global sustainability governance" (The Guardian, 20 June 2012, a). Gro Harlem Brundtland, famous for her 1987 WCED report, also pointed out that "it's not good and it doesn't look good" (ibid.). Apart from a lower number of important heads-of-state compared to what was expected, widespread disappointment regarding the strategy the politicians would finally adopt at Rio+20 predominated the start of the mega-conference. The deal reached by advance negotiators was criticised as too weak to be effective (The Guardian, 20 June 2012, a). For example, Connie Hedegaard, EU climate commissioner, said via twitter that "nobody in that room adopting the text was happy. That's how weak it is" (ibid). JimLeape, head of WWF, said that, "if this text proposed by Brazil is accepted, then the last year of negotiations has been a colossal waste of time (...) you might think Rio+20 was convened as a seminar" (ibid). It seemed that the Brazilian hosts were eager to provide a final conference text before the headsof-states arrived in order to "avoid a repeat of the shambles at the Copenhagen climate conference in 2009" (The Guardian, 20 June, b6).

Having the more or less final text ready before the conference days worried many observers and gave way to fears that "delegates' presence would be reduced to a largely ceremonial role, making – at most – minor tweaks to the agreement" (ibid). The reason for the low expectations at the beginning of the conference regarding the results of Rio+20 seemed to be the lack of clear commitments, timetables, financing or means of monitoring progress. Gro Harlem Brundtland argued that the lack of political commitment may also be due to the current situation of economic and financial crises: "The financial and economic problems that some countries face don't make it easier for them to agree on things that they would have agreed before 2008." (The Guardian, 20 June 2012).

Jeffrey Sachs, director of the Earth Institute at Columbia University, also argued that the timing of the conference was unfortunate in times of debt and budget crises shaking the EU: "Europe has been the great leader of environmental action, but Europe is hardly functioning right now." (New York Times, 18 June). And for President Obama, as argued in the New York Time, it was impossible to go because he had no financial resources to offer and because he would face substantial criticism at home for seeming to be more concerned with global problems than domestic issues – it is a US presidential election year after all (ibid.) Nevertheless, Gro Harlem Brundtland predicted that, like it had been the case in many international negotiations, the final days may produce some surprises: “There are more than 100 leaders coming after all. They are not going to leave with nothing.” (ibid.) As these snippets from the media before the conference show, Rio+20 was being branded as a ‘disappointment’ or ‘failure’ even before it began. The next chapters and sub-chapters will elaborate if this branding was justified by the outcomes Rio+20 delivered.

Rio+20 objectives

The main three objectives of the Ri+20 Conference were:

- to secure renewed political commitment for sustainable development,
- to assess the progress to date and the remaining gaps in the implementation of the outcomes of the major summits on sustainable development, and
- to address new and emerging challenges.

The conference focused mainly on two themes:

I. a green economy in the context of SD and poverty eradication;

II. the institutional framework for SD.

In addition, **seven critical issues**[7] were recognised during the preparatory work for the conference and were given ‘priority attention’: **jobs, energy, cities, food, water, oceans, and disasters**.

Table: Rio+20: seven Critical Issues

Jobs	*Economic action and social policies to create gainful employment are critical for social cohesion and stability. It's also crucial that work is geared to the needs of the natural environment. "Green jobs" are positions in agriculture, industry, services and administration that contribute to preserving or restoring the quality of the environment;*
Energy	*Sustainable energy is needed for strengthening economies, protecting ecosystems and achieving equity. United Nations Secretary-General Ban Ki-moon is leading a Sustainable Energy for All initiative to ensure universal access to modern energy services, improve efficiency and increase use of renewable sources;*
Cities	*At their best, cities have enabled people to advance socially and economically. Many challenges exist to maintaining cities in a way that continues to create jobs and prosperity while not straining land and resources. The challenges cities face can be overcome in ways that allow them to continue to thrive and grow, while improving resource use and reducing pollution and poverty.*
Food	*If done right, agriculture, forestry and fisheries can provide nutritious food for all and generate decent incomes, while supporting people-centred rural development and protecting the environment. (...) soils, freshwater, oceans, forests and biodiversity are being rapidly degraded. Climate change is putting even more pressure on the resources (...). The food and agriculture sector offers key solutions for development, and is central for hunger and poverty eradication.*
Water	*(...) due to bad economics or poor infrastructure, every year millions of people, most of them children, die from diseases associated with inadequate water supply, sanitation and hygiene. Water scarcity, poor water quality and inadequate sanitation negatively impact food security, livelihood choices and educational opportunities for poor families across the world.*
Oceans	*The world's oceans - their temperature, chemistry, currents and life - drive global systems that make the Earth habitable for humankind. (...) Careful management of this essential global resource is a key feature of a sustainable future.*
Disasters	*Disasters caused by earthquakes, floods, droughts, hurricanes, tsunamis and more can have devastating impacts on people, environments and economies. But resilience -- the ability of people and places to withstand these impacts and recover quickly -- remains possible. (...) With a quickening pace of natural disasters taking a greater toll on lives and property, and a higher degree of concentration of human settlements, a smart future means planning ahead and staying alert.*

Theme 1: Green economy in the context of SD and poverty eradication

The Green Economy according to UNEP (2010) is defined as an economy that results in "improved human well-being and social equity, while significantly reducing environmental risks and ecological scarcities"; therefore, an economy that is "low-carbon, resource efficient, and socially inclusive" (UNEP, 2011).

The UNEP's initiative - the Green Economy initiative – has recently reached a climax with the publication of the report 'Towards a Green Economy' (2011), which is defined by UNEP as this initiative's *main output* and that is used as one of the main sources not only for this chapter but also in the official documents that can be found on the Rio+20 official website. In a green economy, "**growth in income and employment are driven by public and private investments that reduce carbon emissions and pollution, enhance energy and resource efficiency, and prevent the loss of biodiversity and ecosystem services**" (ibid.). The "**key aim for a transition to a green economy is to enable economic growth and investment while increasing environmental quality and social inclusiveness**" (ibid.).

Some critical issues with regard to the before mentioned aspects would then be how to: a) to *create the conditions for public and private investments to incorporate broader environmental and social criteria*; b) to *adjust the main indicators of economic performance (such as growth in GDP) in order to account for pollution, resource depletion, declining ecosystem services, and the distributional* consequences of natural capital loss to the poor. The green economy approach is also an "attempt to unite under one banner a broad suite of economic instruments relevant to sustainable development" (UN GA, 2010, p.2)[8]. In the case of the Rio+20 conference, these economic instruments should also be taken into consideration in relation to the global concerns towards poverty eradication.

Additionally, the green economy has to take care of the three pillars of sustainability (environmental protection, social equity and economic development) as being a tool for sustainable development and not its substitute. Nevertheless, apart from achieving a balanced approach on the three pillars of SD poverty eradication should be kept as a primary objective. These two focuses (i.e. green economy and poverty eradication) are not mutually exclusive, but are very much linked. As explained in UNEP's report (2011), in most of the developing countries, the majority of their populations and their livelihoods depends directly on natural resources, while being at the same time especially vulnerable to climate-driven risks (i.e. rising sea levels, coastal erosion, more frequent storms) and to ecological scarcity (i.e. water scarcity, access to clean water and basic sanitation). While a green economy – per se – will not address all poverty issues, a 'pro-poor' orientation in its initiatives and a focus towards "finding ways to protect global ecosystems, reduce the risks of global climate change, improve energy security, and simultaneously improve the livelihoods of the poor" (p. 20) will contribute to SD.

In order to come up to the interests of environmental protection and development by developing and developed countries, respectively, UNEP (2011) argues that a green economy could provide a "development path that reduces carbon dependency, promotes resource and energy efficiency andlessens environmental degradation" (p.17) and, at the same time, "reconciling the competing economic development aspirations of rich and poor countries in a world economy that is facing increasing climate change, energy insecurity and ecological scarcity" (p.16).

Another important source in this context is the work done by the Organisation for Economic Cooperation and Development (OECD) on the similar concept of 'Green growth' that is defined as:

> "fostering economic growth and development while ensuring that natural assets continue to provide the resources and environmental services on which our well-being relies. To do this it must catalyse investment and innovation which will underpin sustained growth and give rise to new economic opportunities." (OECD, 2011, p.9)

In fact, the OECD Green Growth Strategy is described as developing "a clear and focused agenda for delivering on a number of Rio's key aspirations [whilst] not been conceived as a replacement for sustainable development, but rather (…) a subset of it" (OECD, 2011, p.11).

For a critical perspective on the Green Economy please see the Draft ESDN Policy Brief for the Rio+20 Side event on "National Sustainable Development Strategies – What Future Role with Respect to Green Economy?

Theme 2: Institutional framework for sustainable development (IFSD)

The second theme of the conference builds on the mandate of the Johannesburg Plan of Implementation (JPOI) that was the outcome document of the World Summit on Sustainable

Development (WSSD) held in Johannesburg in 2002. Chapter XI of the JPOI addresses the **necessity to strengthen the Institutional Framework for Sustainable Development (IFSD)**.

Precisely, in paragraph 137 the JPOI affirms:

> "*An effective institutional framework for sustainable development at all levels is key to the full implementation of Agenda 21, the follow-up to the outcomes of the World Summit on Sustainable Development and meeting emerging sustainable development challenges.*
>
> *(...) It should be responsive to the needs of all countries, taking into account the specific needs of developing countries including the means of implementation.*
>
> *It should lead to the strengthening of international bodies and organizations dealing with sustainable development, while respecting their existing mandates, as well as to the strengthening of relevant regional, national and local institutions.*"

Whilst, the subsequent paragraph 138 in the JPOI explains how "Good governance is essential for sustainable development", paragraph 139 states the need to take a number of measures to "**strengthen institutional arrangements on sustainable development, at all levels**" in order to achieve a list of nine objectives:

1. Strengthening commitments to sustainable development;

2. Integration of the economic, social and environmental dimensions of sustainable development in a balanced manner;

3. Strengthening of the implementation of Agenda 21, including through the mobilization of financial and technological resources, as well as capacity-building programmes, particularly for developing countries;

4. Strengthening coherence, coordination and monitoring;

5. Promoting the rule of law and strengthening of governmental institutions;

6. Increasing effectiveness and efficiency through limiting overlap and duplication of activities of international organizations, within and outside the United Nations system, based on their mandates and comparative advantages;

7. Enhancing participation and effective involvement of civil society and other relevant stakeholders in the

8. implementation of Agenda 21, as well as promoting transparency and broad public participation;

9. Strengthening capacities for sustainable development at all levels, including the local level, in particular those of developing countries;

10. Strengthening international cooperation aimed at reinforcing the implementation of Agenda 21 and the outcomes of the Summit.

In addition, the JPOI identifies sustainable development as an overarching goal for institutions at the national, regional and international levels. In this discussion, the Nairobi-Helsinki Outcome gained great importance. Building on the work of the Belgrade process, and adopted by the Consultative Group of Ministers or High-level Representatives on International Environmental Governance, the 'outcome' identified five options for broader reform that will be discussed in Rio, namely:

a) Enhancing UNEP;

b) Establishing a new umbrella organization for sustainable development;

c) Creating a specialized agency such as a world environment organization;

d) Introducing possible reforms to ECOSOC and the CSD; and

e) Enhanced institutional reforms and streamlining of present structures.

These options were then analysed again in Solo (Indonesia, 19-21 July 2011) by the High Level Dialogue on Institutional Framework for Sustainable Development (IFSD) that produced progress in the discussions as part of the Rio+20 preparations, which is summarized as follows:

1. **Linkages among the three pillars: three main options received most consensuses:**

a. strengthening CSD,

b. adjusting the mandate of ECOSOC, and

c. establishing a Sustainable Development Council, that received the most "heightened interest (…) by all groups of countries";

2. **Strengthening UNEP**: again consensus was registered and, especially, the "willingness by all groups of countries to explore the question of a specialized agency status";

3. **Delivering as one**: while recognising "the need to balance a top down approach to sustainable development with a bottom up approach based on the expressed needs of countries", special attention was devoted "to enhance the coherence of international support to national sustainable development plans";

4. **Science-Policy Interface**: reaffirming the "need to strengthen the linkages between science and policy", the "idea of an Inter-governmental Panel on Sustainable Development (IPSD), along the lines of the IPCC" was put forward;

5. **Financing**: discussions were mainly directed to:
a. a dedicated fund for sustainable development,
b. an adequate and additional finance needed for implementation and,
c. the use of innovative sources of financing to complement ODA.

6. **Sustainable Development Goals (SDGs)**: although not easy to converge upon, it was felt that future agreement on SDGs development might be found;

7. **National Level Institutions**: consensus was found on the need for "more systematic mainstreaming of a sustainable development perspective into the work of the key economic ministries";

8. **Involving non-State Actors**: convergence was found on the need to support "the engagement between governments and non-State actors (…) as a means of strengthening both decision making and implementation of sustainable development".

The process of developing an international set of Sustainable Development Goals (SDGs)

In the course of preparing the Rio+20 conference, growing support has been given to an elaboration of global Sustainable Development Goals (SDGs) by 2015 'based on Agenda 21 and Johannesburg Plan of Implementation, fully respect all Rio Principles…' and 'for pursuing focused and coherent action on sustainable development' (UN Final 'Draft' of the outcome document: para. 246, 2012), and to develop mechanisms for their monitoring and reporting, as well as to develop indicators complementing GDP in measuring well-being and integrating economic, social, and environmental dimensions. The initial debate on SDGs has been backed and further elaborated in the outcome document of the UN Deparment of Public Information/NGO Conference (Bonn, Germany, September 2011) and by the UN Secretary General's High level Panel on Global Sustainability10. Furthermore, gathered wider political attention at the subsequent Informal and Inter-sessional negotiations convened to elaborate the Zero draft document and are featured prominently in the Final 'Draft' presented to the world leader11 of the Rio+20 Outcome Document.

A summary of the various consultations on SDGs on the way to elaborate the Zero draft document in terms of its objectives, characteristics, and the scope of SDGs is shown below in table 1 (references to the Final outcome document are indicated in parenthesis).

Table: Objectives, Characteristics and Scope of SDGs

Objectives	Address broader challenges threatening sustainable development
	Reaffirm the past political commitments of all actors and ensure tangible actions towards sustainable development (Final outcome document, para. 246, 2012)
Characteristics	Action-oriented (para. 247, 2012)
	Complementary to MDGs (para. 246, 2012)
	Strongly linked to Agenda 21 and JPoI (para. 246, 2012)
	Universal in application, but allowing for national and regional circumstances and respective capabilities (para. 247, 2012)
	Voluntary application, in keeping with national realities, priorities, and capabilities (para. 247, 2012)
Scope	Poverty eradication as an overarching goal
	Address economic, social, and environmental dimensions of sustainable development
	Enable articulation of the nexus between the different issue areas covered by the SDGs
	Time bound and measurable, with targets and indicators
	Few in number and easy to communicate and understand (para. 247, 2012)

The development of SDGs could assist in focusing the broad international sustainable development agenda at a practical level, and in the case of the Millennium Development Goal (MDG) framework, could act as extension of the original framework in its post-2015 period. Furthermore, newly formed or added SDGs might benefit from wide political commitment, experience and support through MDGs because of the successes of the MDGs in:

> 1) Rallying public, private and political support for global poverty reduction, 2) providing an effective tool to stimulate the production of new poverty-related data and additional aid commitments, and facilitating greater coordination of international development efforts between nation states and other development actors.

However, as the current MDG framework does not fully cover emerging or urgent issues such as climate change, energy security, resilience or disaster preparedness, an upcoming set of SDGs could address shortcomings and challenges of the MDGs and broaden their goals to reflect other SD objectives. In this regard, Lingán et al. (2012) confirm that, for example, the MDGs have not assisted developing countries in governance related issues such as development of robust government institutions, social welfare systems and an enabling environment for civil society and, thus, left them particularly vulnerable to emerging global development challenges such as climate change. Moreover, an elaboration of SDGs would need to take into account that the UN process on MDGs is still on-going with regard to implementation and review of the Millennium Development Goals and the post-2015 UN Development Agenda.

The United Nations Secretary-General's High Level Panel on Global Sustainability Report "Resilient people, resilient planet" (2012) recommends to governments to agree on the development of a set of key universal sustainable development goals, covering all three dimensions of sustainable development as well as their interconnections. So far, several options for Rio+20 deliverables have been articulated during on-going informal consultations on SDGs, organised by the Government of Columbia.

The EU Position

Many actors, especially from the European Parliament and the Council of the European Union, advocated for a strong role of the European Union at the Rio+20 Conference. Accordingly, in 2011, the European Commission declared to be "determined to help make Rio+20 a success". The European Union committed itself to play an active and constructive role in order to achieve global action on how to liftpeople out of poverty and how to use resources better to ensure prosperity also for future generations. *On behalf of the European Commission, the Directorate-General for Environment leaded the negotiations at the RIO +20 Conference.*

A number of official documents delineate where the EU stands. In this regards, particularly relevant are the following three official documents:

- the European Commission's communication "Rio+20: towards the green economy and better governance";
- the Council of the European Union's Council conclusions "Rio+20: Pathways to a Sustainable Future";
- the European Council's conclusions "1/2 March 2012".

The Communication"Rio+20: towards the green economy and better governance" sets out the Commission's initial views as part of the preparatory process that leads to the Rio+20 Conference where the EU seeks "tangible actions (…) in enabling the transition to the green economy and better governance" (EC, 2011).

The European Commission expressed very clearly its opinion on the importance of the Rio+20 conference. On the one hand, it is pointed out that "Rio+20 offers a unique opportunity for our mutually interdependent world to secure renewed political commitment for sustainable development [while assessing] progress made and address[ing] implementation gaps and emerging challenges" (EC, 2011). On the other hand, the Commission highlighted that "Rio+20 can mark the start of an accelerated and profound, world-wide transition towards a green economy – an economy that generates growth, creates jobs and eradicates poverty by investing in and preserving the natural capital offers upon which the long-term survival of our planet depends. It can also launch the needed reform of international sustainable development governance" (EC, 2011).

Therefore, in the Communication, the EC suggested four crucial points to be followed to boost sustainable development at the Rio+20 Conference:

1. broad political "rallying call" with a shared, ambitious vision and goals;

2. a set of specific actions at international, regional and national level - mapped out as a "Green Economy Roadmap";

3. "toolbox" of policy approaches and best practice examples to be used to reach agreed objectives;

4. a mechanism to promote and monitor overall progress. A particular focus of these suggestions seems to be on the necessity of an internationally agreed **Green Economy Roadmap** in order to "guarantee continued commitment beyond Rio+20, ensuring that the agreed vision and goals will be followed through in a systematic manner (…) with milestones, indicators and targets, as well as mechanisms to monitor overall progress". Essential seems therefore the establishment of strategies for greening the economy as part of the overall economic and development policies and plans of countries while, at the same time, the Green Economy Roadmap needs also to include actions at global and regional level.

Specific actions suggested by the European Commission towards Rio+20

1. Action on resources, materials and natural capital

a. promote sustainable water and establish an international partnerships on water;

b. increase energy access, energy security and promote renewable energy and energy efficiency;

c. strengthen protection of the marine environment and oceans;

d. promote sustainable agriculture, land-use and food security and, to make the consumption and production of food commodities more sustainable, establish international partnerships on food commodities;

e. promote sustainable forest management and combat deforestation;

f. establish a more robust and coherent international regime on chemicals and hazardous substances;

g. put in motion a mechanism for global science and research cooperation on societal challenges of global importance (e.g. resource constraints, climate change, oceans).

2. Providing economic instruments and financing and investing in human capital

a. develop domestic and regional carbon emission trading schemes;

b. identify and phase out environmentally harmful subsidies, accompanied by targets and deadlines;

c. consolidate and strengthen existing financing strategies and facilities, or establish new public-private financing schemes;

d. establish green skills training programmes in priority areas such as energy, agriculture, construction, natural resource management, waste and recycling.

3. Improving governance

a. strengthen sustainable development governance within the UN (i.e. reinforce the role of the UN ECOSOC; upgrade the UNCSD to a permanent body; give more emphasis on SD in all relevant UN bodies; …);

b. UNEP needs to be reinforced, for example by transforming it into a UN Specialized Agency (such as ILO);

c. Accelerate the work on streamlining and reinforcing the MEA system, as part of the strengthening of international environmental governance;

d. Reinforce capacity building for the environment within the UN strengthening environmental expertise and green economy awareness within UN country teams to promote mainstreaming in country programmes;

e. Strengthen the capacity to monitor the global environment;

f. strengthen the engagement of the private sector.

On the 9 Match 2012, the Council of the European Union's in its Council conclusions, entitled "Rio+20: Pathways to a Sustainable Future", reiterated its support to the road drawn by the European Commission in June 2011.

Most recently (14-15 June 2012), ministers of the EU and African, Caribbean and Pacific countries agreed on a joint declaration ahead of the Rio+20 Conference where both parties pledged their commitment "to working constructively during the Conference to ensure an ambitious and action oriented outcome". They reaffirmed their "strong belief that the United Nations Conference on Sustainable Development is a unique opportunity to ensure a renewed political and international commitment for advancing the sustainable development agenda based on the assessment of progress, in the fulfilment of those commitments made to date as well as on new and emerging challenges".

They also "acknowledge that a global transition to an all-inclusive green economy could contribute to achieving sustainable development through inter alia, poverty alleviation, increased employment, improved land management, forest conservation, enhanced food security, improved management of water resources, resource efficiency and increased access to sustainable energy, while integrating and building on the value of natural capital and thus sustaining environmental resources and ecological services essential for development".

The Preparatory Process and the Negotiations Pre –Rio

On the 24 December 2009, the UN General Assembly (UNGA 64) adopted Resolution 64/236 and agreed to convene the UNCSD in Rio, Brazil.[14] In preparation for the Conference, and with the purpose of discussing substantive and procedural issues, three Preparatory Committee (PrepCom) meetings were held in the context of the UN. This decision was taken in resolutions 64/236 and 65/152. Also an inclusive preparatory process was carried out with the aim of involving various stakeholders and at different levels.

The first meeting was held between 16 and 18 May 2010 in New York and discussed the agreed substantive themes of the Conference, addressed pending procedural matters, and elected the Bureau. Elected by Member States, **the 10-member Bureau** was formed by 2 representatives from each region, and Brazil as ex-officio member to steer the preparatory process and decide on the roadmap and organization of work of the preparatory process. On 7-8 March 2011, in New York, **the second preparatory meeting** was held with the objective to discuss further the substantive themes of the Conference. **A third meeting took** place one week prior to the Conference itself, between 13 and 15 June 2012. Additionally, three inter-sessional meetings took place: **one in January 2010, the second one** in December 2011 and **the third one in** March 2012, in New York. During the second half of 2011, a series of regional and sub-regional meetings were also held to prepare inputs for the UNCSD preparatory process.

The Initial discussions of the 'Zero Draft' of the outcome document started in a meeting that convened at the UN Headquarters from 25-27 January 2012 in New York. Following this meeting, a first "Informal Informal" consultations meeting was held in March 2012 together with the third intersessional meeting. In this meeting, "delegates engaged in lengthy discussions on the text, proposing amendments and responding to other delegations' suggestions. (…) most sections of the text had been reviewed and discussed more than once, with the text expanding to more than 200 pages" (ENB, 2012) [15]. A second round of "Informal Informal" consultations was held from 23 April to 4 May 2012. In this meeting, delegates "agreed ad referendum to 21 out of 420 paragraphs in the text, and so the Bureau decided to hold an additional negotiating session prior to the UNCSD" (ibid.).

From 29 May to 2 June 2012, a third round of "informal informal" consultations on the draft outcome was again held in New York. "Delegates discussed the 80-page revised draft text produced by the CoChairs, [agreeing] 70 paragraphs (…) ad referendum, with 259 containing bracketed text" (ibid.). The third PrepCom meeting did not produced a complete and agreed 'draft' of the outcome document. It concluded at 12:16 am on Saturday, 16 June 2012, following a full day of negotiations in multiple "splinter" groups and informal consultations. Finally, Brazil was invited in the role of organiser to conduct "pre- conference informal consultations led by the host country" since only 116 paragraphs were agreed ad referendum while 199 were yet to be agreed.
Finally, on Tuesday 19th of June, during a plenary meeting of the Pre-Conference Informal Consultations, delegates agreed to the outcome **document *ad referendum***.

The Final Outcome Document

The final 'draft' outcome document16 – The future we want – was published on 19 June 2012 during a plenary meeting of the Pre-Conference Informal Consultations. During the conference from 20-22 June, this document was then handed over to and discussed by the head of states and their representatives and, finally, became the "The Future We Want - Outcome Document"17 (referred to here as "Final outcome document”). More specifically, the document was based on the so-called ‘zero draft’, a draft of the outcome document that was submitted by the co-Chairs on behalf of the Bureau in accordance with the decision taken in the second preparatory committee meeting, also known as Prepcom 2. The objective was to present a Zero Draft to be considered by Member States and other stakeholders no later than early January 2012. Therefore, the Zero Draft was published on 10 January 2012; it consisted of 19 pages with 128 paragraphs divided in five sections, with a table of contents on the first two pages.

The final outcome document of Rio+20 is very different from the Zero Draft. In fact, the final outcome document consists of 53 pages instead of 19 and 283 paragraphs instead of 128. Additionally, the structure changed notably and we show the differences in red in the box-text below. The new structure comprehends six main sections:

I. Our Common vision

II. Renewing Political Commitment

III. Green Economy in the context of sustainable development and poverty eradication

IV. Institutional Framework for Sustainable Development

V. Framework for action and follow-up

VI. Means of Implementation

Our Common Vision

The first section delineates the vision of the document in 13 paragraphs that mainly leverages on a globally renewed commitment toward sustainable development while “ensuring the “promotion of economically, socially and environmentally sustainable future for our planet and for present and future generations” (UNCSD, 2012). Poverty eradication is then recognized as the greatest global challenge and as an “indispensable requirement for sustainable development”.
In the fourth paragraph, it is especially well described what is intended for sustainable development, its objectives and its requirements:

> *We recognize that* ***poverty eradication, changing unsustainable and promoting sustainable patterns of consumption and production and protecting and managing the natural resource base of economic and social development*** *are the overarching objectives of and essential requirements for sustainable development. We also reaffirm the need to achieve sustainable development by promoting sustained, inclusive and equitable economic growth, creating greater opportunities for all, reducing inequalities, raising basic standards of living, fostering equitable social development and inclusion, and promoting integrated and sustainable management of natural resources and ecosystems that supports, inter alia, economic, social and human development while facilitating ecosystem conservation, regeneration and restoration and resilience in the face of new and emerging challenges.*

In the second part of the fourth paragraph, it is interesting to notice how strong the European Union’s vision in this formulation apparently was (see also the European Commission’s 2011 Communication); this is especially visible in the passage “promoting *sustained, inclusive and equitable economic growth”.* Although very much focused on solving the poverty issue and sustainable development through the ‘usual recipe’ of economic growth, it is also possible to highlight the role of natural resources and of ecosystems conservation in this text, suggesting progress for the environmental pillar of SD. Also social justice and equity is present (especially in paragraph 6) and words such *freedom*, *peace*, *democracy* and *participation* in decision-making, are stressed in the text.

Renewing Political Commitment

Section two, **Renewing Political Commitment**, consists of 42 paragraphs divided in 3 subsections. First, it mainly reaffirms the Rio 1992 principles and past action plans. Secondly, it is suggested to undertake an assessment of progresses made and gaps that have remained since the 1992 Rio Declaration. In this context, especially four urgent problems seem to be still far away from the solution:

I. Poverty (par.21, 23),

II. Unemployment (par.24),

III. Climate Change (par.25), and

IV. the relationship between people and ecosystems (and particularly considering the poor and their livelihoods) (par.30).

Especially interesting from the perspective of the last two points just mentioned - and therefore on the role of ecosystems - is paragraph 39 that recognises not only the importance of nature in itself – and "Mother Earth" - but also stressed the recognition by some countries of the "rights of nature":

> "39. We recognize that the planet Earth and its ecosystems are our home and that Mother Earth is a common expression in a number of countries and regions and we note that some countries recognize the rights of nature in the context of the promotion of sustainable development. We are convinced that in order to achieve a just balance among the economic, social and environment needs of present and future generations, it is necessary to promote harmony with nature."

The third subsection, in its 14 paragraphs, stresses the necessity of "engaging major groups and other stakeholders" and highlighted especially that "broad public participation and access to information and judicial and administrative proceedings are essential to the promotion of sustainable development" that also requires involvement and active participation (par.43).

Green economy in the context of SD and poverty eradication

The first of the two Conference's themes is then treated in section three, which is titled accordingly Green Economy in the context of sustainable development and poverty eradication and is tackled in 19 paragraphs.

While addressing the role of the green economy as an important tool for SD and poverty eradication, it is, nonetheless, undoubtedly reaffirmed that sustainable development is humankind's overarching goal.

Accordingly, no critiques can be advanced on the possibility of a substitution of the green economy as a global goal for humanity.

Paragraph 58 appeared to be a crucial passage because it somehow explained how this green economy should work. In this regards, it is again possible to notice a lower importance of the environmental pillars, which it is not considered in this key paragraph. Environment is only touched in the following paragraphs 60 and 61 where one could have expected its inclusion in the ‘functioning’ of the green economy.

In paragraph 59, it is suggested that countries, when implementing green economy policies, they can choose an appropriate approach “in accordance with national sustainable development plans, strategies and priorities”.

Another critique is here advised: it is suggested that, especially reading between the lines of paragraph 56, 58 (letters b, e and h), 59 and 74, real commitments and practical decisions are evaded and, moreover, there seems to be a very visible effort to underplay the possibility to have similar targets and similar green economy strategies. For instance: par. 59 says “(…) each country can choose an appropriate approach in accordance with national sustainable development plans, strategies and priorities”. Also indicative is par. 58 (h) that addresses the possibility of posing trade closures for environmental matters; in fact, it affirms, “avoid unilateral actions to deal with environmental challenges outside the jurisdiction of the importing country”.

A critical perspective on the relationship between the Green Economy and sustainable development can be found in the background paper of the Rio+20 side-event, the ESDN organised in cooperation with Switzerland and Liechtenstein.

Unit-11

2.3 Montreal Protocol

The Montreal Protocol on Substances that Deplete the Ozone Layer (a protocol to the Vienna Convention for the Protection of the Ozone Layer) is an international treaty designed to protect the ozone layer by phasing out the production of numerous substances that are responsible for ozone depletion. It was agreed on 26 August 1987, and entered into force on 16 September 1989, following a first meeting in Helsinki, May 1989. Since then, it has undergone nine revisions, in 1990 (London), 1991 (Nairobi), 1992 (Copenhagen), 1993 (Bangkok), 1995 (Vienna), 1997 (Montreal), 1998 (Australia), 1999 (Beijing) and 2016 (Kigali).

As a result of the international agreement, the ozone hole in Antarctica is slowly recovering. Climate projections indicate that the ozone layer will return to 1980 levels between 2050 and 2070.[5][6] Due to its widespread adoption and implementation it has been hailed as an example of exceptional international co-operation, with Kofi Annan quoted as saying that "perhaps the single most successful international agreement to date has been the Montreal Protocol".[7][8] In comparison, effective burden sharing and solution proposals mitigating regional conflicts of interest have been among the success factors for the ozone depletion challenge, where global regulation based on the Kyoto Protocol has failed to do so.[9] In this case of the ozone depletion challenge, there was global regulation already being installed before a scientific consensus was established. Also, overall public opinion was convinced of possible imminent risks.

The two ozone treaties have been ratified by 197 parties (196 states and the European Union), making them the first universally ratified treaties in United Nations history. These truly universal treaties have also been remarkable in the expedience of the policy-making process at the global scale, where only 14 years lapsed between a basic scientific research discovery (1973) and the international agreement signed (1985 and 1987).

Chlorofluorocarbons (CFCs) Phase-out Management Plan

The stated purpose of the treaty is that the signatory states

> "Recognizing that worldwide emissions of certain substances can significantly deplete and otherwise modify the ozone layer in a manner that is likely to result in adverse effects on human health and the environment. Determined to protect the ozone layer by taking precautionary measures to control equitably total global emissions of substances that deplete it with the ultimate objective of their elimination on the basis of developments in scientific knowledge"

"Acknowledging that special provision is required to meet the needs of developing countries" shall accept a series of stepped limits on CFC use and production, including:

from 1991 to 1992 its levels of consumption and production of the controlled substances in Group I of Annex A do not exceed 150 percent of its calculated levels of production and consumption of those substances in 1986;

from 1994 its calculated level of consumption and production of the controlled substances in Group I of Annex A does not exceed, annually, twenty-five percent of its calculated level of consumption and production in 1986.

from 1996 its calculated level of consumption and production of the controlled substances in Group I of Annex A does not exceed zero.

There was a faster phase-out of halon-1211, -2402, -1301, There was a slower phase-out (to zero by 2010) of other substances (halon 1211, 1301, 2402; CFCs 13, 111, 112, etc.)[contradictory] and some chemicals were given individual attention (Carbon tetrachloride; 1,1,1-trichloroethane). The phasing-out of the less damaging HCFCs only began in 1996 and will go on until a complete phasing-out is achieved by 2030.

There were a few exceptions for "essential uses" where no acceptable substitutes were initially found (for example, in the past metered dose inhalers commonly used to treat asthma and chronic obstructive pulmonary disease were exempt) or Halon fire suppression systems used in submarines and aircraft (but not in general industry).

The substances in Group I of Annex A are:

- CFCl3 (CFC-11)
- CF2Cl2 (CFC-12)
- C2F3Cl3 (CFC-113)
- C2F4Cl2(CFC-114)
- C2F5Cl (CFC-115)

The provisions of the Protocol include the requirement that the Parties to the Protocol base their future decisions on the current scientific, environmental, technical, and economic information that is assessed through panels drawn from the worldwide expert communities. To provide that input to the decision-making process, advances in understanding on these topics were assessed in 1989, 1991, 1994, 1998 and 2002 in a series of reports entitled Scientific assessment of ozone depletion, by the Scientific Assessment Panel (SAP).

In 1990 a Technology and Economic Assessment Panel was also established as the technology and economics advisory body to the Montreal Protocol Parties. Technology and Economic Assessment Panel (TEAP) provides, at the request of Parties, technical information related to the alternative technologies that have been investigated and employed to make it possible to virtually eliminate use of Ozone Depleting Substances (such as CFCs and Halons), that harm the ozone layer. The TEAP is also tasked by the Parties every year to assess and evaluate various technical issues including evaluating nominations for essential use exemptions for CFCs and halons, and nominations for critical use exemptions for methyl bromide. TEAP's annual reports are a basis for the Parties' informed decision-making.

Numerous reports have been published by various inter-governmental, governmental and non-governmental organizations to catalogue and assess alternatives to the ozone depleting substances, since the substances have been used in various technical sectors, like in refrigeration, air conditioning, flexible and rigid foam, fire protection, aerospace, electronics, agriculture, and laboratory measurements.]

Hydrochlorofluorocarbons (HCFCs)

Phase-out Management Plan (HPMP)

Under the Montreal Protocol on Substances that Deplete the Ozone Layer, especially Executive Committee (ExCom) 53/37 and ExCom 54/39, Parties to this Protocol agreed to set year 2013 as the time to freeze the consumption and production of HCFCs for developing countries. For developed countries, reduction of HCFC consumption and production began in 2004 and 2010, respectively, with 100% reduction set for 2020. Developing countries agreed to start reducing its consumption and production of HCFCs by 2015, with 100% reduction set for 2030.[21]

Hydrochlorofluorocarbons, commonly known as HCFCs, are a group of man-made compounds containing hydrogen, chlorine, fluorine and carbon. They are not found anywhere in nature. HCFC production began to take off after countries agreed to phase out the use of CFCs in the 1980s, which were found to be destroying the ozone layer. Like CFCs, HCFCs are used for refrigeration, aerosol propellants, foam manufacture and air conditioning. Unlike the CFCs, however, most HCFCs are broken down in the lowest part of the atmosphere and pose a much smaller risk to the ozone layer. Nevertheless, HCFCs are very potent greenhouse gases, despite their very low atmospheric concentrations, measured in parts per trillion (million million).

The HCFCs are transitional CFCs replacements, used as refrigerants, solvents, blowing agents for plastic foam manufacture, and fire extinguishers. In terms of ozone depletion potential (ODP), in comparison to CFCs that have ODP 0.6 – 1.0, these HCFCs have lower ODPs (0.01 – 0.5). In terms of global warming potential (GWP), in comparison to CFCs that have GWP 4,680 – 10,720, HCFCs have lower GWPs (76 – 2,270).

Hydrofluorocarbons (HFCs)

On January 1, 2019 the Kigali Amendment to the Montreal Protocol came into force. Under the Kigali Amendment countries promised to reduce the use of hydrofluorocarbons (HFCs) by more than 80% over the next 30 years. By December 27, 2018, 65 countries had ratified the Amendment.

Produced mostly in developed countries, hydrofluorocarbons (HFCs) replaced CFCs and HCFCs. HFCs pose no harm to the ozone layer because, unlike CFCs and HCFCs, they do not contain chlorine. They are, however, greenhouse gases, with a high global warming potential (GWP), comparable to that of CFCs and HCFCs. Thus, in 2009, a study calculated that a fast phasedown of high-GWP HFCs could potentially prevent the equivalent of up to 8.8 Gt CO2-eq per year in emissions by 2050. A proposed phasedown of HFCs was hence projected to avoid up to 0.5C of warming by 2100 under the high-HFC growth scenario, and up to 0.35C under the low-HFC growth scenario. Recognizing the opportunity presented for fast and effective phasing down of HFCs through the Montreal Protocol, starting in 2009 the Federated States of Micronesia proposed an amendment to phase down high-GWP HFCs, with the U.S., Canada, and Mexico following with a similar proposal in 2010.

After seven years of negotiations, in October 2016 at the 28th Meeting of the Parties to the Montreal Protocol in Kigali, the Parties to the Montreal Protocol adopted the Kigali Amendment whereby the Parties agreed to phasedown HFCs under the Montreal Protocol. The amendment to the legally- binding Montreal Protocol will ensure that industrialised countries bring down their HFC production and consumption by at least 85 per cent compared to their annual average values in the period 2011-2013. A group of developing countries including China, Brazil and South Africa are mandated to reduce their HFC use by 85 per cent of their average value in 2020-22 by the year 2045. India and some other developing countries — Iran, Iraq, Pakistan, and some oil economies like Saudi Arabia and Kuwait — will cut down their HFCs by 85 per cent of their values in 2024-26 by the year 2047.

On 17 November 2017, ahead of the 29th Meeting of the Parties of the Montreal Protocol, Sweden became the 20th Party to ratify the Kigali Amendment, pushing the Amendment over its ratification threshold ensuring that the Amendment would enter into force 1 January 2019.

History

In 1973, the chemists Frank Sherwood Rowland and Mario Molina, who were then at the University of California, Irvine, began studying the impacts of CFCs in the Earth's atmosphere. They discovered that CFC molecules were stable enough to remain in the atmosphere until they got up into the middle of the stratosphere where they would finally (after an average of 50–100 years for two common CFCs) be broken down by ultraviolet radiation releasing a chlorine atom. Rowland and Molina then proposed that these chlorine atoms might be expected to cause the breakdown of large amounts of ozone (O3) in the stratosphere. Their argument was based upon an analogy to contemporary work by Paul J. Crutzen and Harold Johnston, which had shown that nitric oxide (NO) could catalyze the destruction of ozone. (Several other scientists, including Ralph Cicerone, Richard Stolarski, Michael McElroy, and Steven Wofsy had independently proposed that chlorine could catalyze ozone loss, but none had realized that CFCs were a potentially large source of chlorine.) Crutzen, Molina and Rowland were awarded the 1995 Nobel Prize for Chemistry for their work on this problem.

The environmental consequence of this discovery was that, since stratospheric ozone absorbs most of the ultraviolet-B (UV-B) radiation reaching the surface of the planet, depletion of the ozone layer by CFCs would lead to an increase in UV-B radiation at the surface, resulting in an increase in skin cancer and other impacts such as damage to crops and to marine phytoplankton.

But the Rowland-Molina hypothesis was strongly disputed by representatives of the aerosol and halocarbon industries. The chair of the board of DuPont was quoted as saying that ozone depletion theory is "a science fiction tale...a load of rubbish...utter nonsense". Robert Abplanalp, the president of Precision Valve Corporation (and inventor of the first practical aerosol spray can valve), wrote to the Chancellor of UC Irvine to complain about Rowland's public statements (Roan, p. 56.)

After publishing their pivotal paper in June 1974, Rowland and Molina testified at a hearing before the U.S. House of Representatives in December 1974. As a result, significant funding was made available to study various aspects of the problem and to confirm the initial findings. In 1976, the U.S.

National Academy of Sciences (NAS) released a report that confirmed the scientific credibility of the ozone depletion hypothesis. NAS continued to publish assessments of related science for the next decade.

Then, in 1985, British Antarctic Survey scientists Joe Farman, Brian Gardiner and Jonathan Shanklin published results of abnormally low ozone concentrations above Halley Bay near the South Pole. They speculated that this was connected to increased levels of CFCs in the atmosphere. It took several other attempts to establish the Antarctic losses as real and significant, especially after NASA had retrieved matching data from its satellite recordings. The impact of these studies, the metaphor 'ozone hole', and the colourful visual representation in a time lapse animation proved shocking enough for negotiators in Montreal, Canada to take the issue seriously.

Also in 1985, 20 nations, including most of the major CFC producers, signed the Vienna Convention, which established a framework for negotiating international regulations on ozone-depleting substances. After the discovery of the ozone hole by SAGE 2 it only took 18 months to reach a binding agreement in Montreal, Canada.

But the CFC industry did not give up that easily. As late as 1986, the Alliance for Responsible CFC Policy (an association representing the CFC industry founded by DuPont) was still arguing that the science was too uncertain to justify any action. In 1987, DuPont testified before the US Congress that "We believe there is no imminent crisis that demands unilateral regulation." And even in March 1988, Du Pont Chair Richard E. Heckert would write in a letter to the United States Senate, "we will not produce a product unless it can be made, used, handled and disposed of safely and consistent with appropriate safety, health and environmental quality criteria. At the moment, scientific evidence does not point to the need for dramatic CFC emission reductions. There is no available measure of the contribution of CFCs to any observed ozone change..."

Multilateral Fund

The main objective of the Multilateral Fund for the Implementation of the Montreal Protocol is to assist developing country parties to the Montreal Protocol whose annual per capita consumption and production of ozone depleting substances (ODS) is less than 0.3 kg to comply with the control measures of the Protocol. Currently, 147 of the 196 Parties to the Montreal Protocol meet these criteria (they are referred to as Article 5 countries).

It embodies the principle agreed at the United Nations Conference on Environment and Development in 1992 that countries have a common but differentiated responsibility to protect and manage the global commons.

The Fund is managed by an Executive Committee with an equal representation of seven industrialized and seven Article 5 countries, which are elected annually by a Meeting of the Parties. The Committee reports annually to the Meeting of the Parties on its operations. The work of the Multilateral Fund on the ground in developing countries is carried out by four Implementing Agencies, which have contractual agreements with the Executive Committee:

- United Nations Environment Programme (UNEP), through its OzonAction Programme.
- United Nations Development Programme (UNDP).
- United Nations Industrial Development Organization (UNIDO).
- World Bank.

Up to 20 per cent of the contributions of contributing parties can also be delivered through their bilateral agencies in the form of eligible projects and activities.

The fund is replenished on a three-year basis by the donors. Pledges amount to US$3.1 billion over the period 1991 to 2005. Funds are used, for example, to finance the conversion of existing manufacturing processes, train personnel, pay royalties and patent rights on new technologies, and establish national ozone offices.

Parties

As of 23 June 2015, all countries in the United Nations, the Cook Islands, Holy See, Niue as well as the European Union have ratified the original Montreal Protocol (see external link below), with South Sudan being the last country to ratify the agreement, bringing the total to 197. These countries have also ratified the London, Copenhagen, Montreal, and Beijing amendments.

Effect

Ozone-depleting gas trends

Since the Montreal Protocol came into effect, the atmospheric concentrations of the most important chlorofluorocarbons and related chlorinated hydrocarbons have either leveled off or decreased.[36] Halon concentrations have continued to increase, as the halons presently stored in fire extinguishers are released, but their rate of increase has slowed and their abundances are expected to begin to decline by about 2020. Also, the concentration of the HCFCs increased drastically at least partly because of many uses (e.g. used as solvents or refrigerating agents) CFCs were substituted with HCFCs.

While there have been reports of attempts by individuals to circumvent the ban, e.g. by smuggling CFCs from undeveloped to developed nations, the overall level of compliance has been high. Statistical analysis from 2010 shows a clear positive signal from the Montreal Protocol to the stratospheric ozone. In consequence, the Montreal Protocol has often been called the most successful international environmental agreement to date. In a 2001 report, NASA found the ozone thinning over Antarctica had remained the same thickness for the previous three years, however in 2003 the ozone hole grew to its second largest size. The most recent (2006) scientific evaluation of the effects of the Montreal Protocol states, "The Montreal Protocol is working: There is clear evidence of a decrease in the atmospheric burden of ozone-depleting substances and some early signs of stratospheric ozone recovery." However, a more recent study seems to point to a relative increase in CFCs due to an unknown source.

Reported in 1997, significant production of CFCs occurred in Russia for sale on the black market to the EU throughout the 90s. Related US production and consumption was enabled by fraudulent reporting due to poor enforcement mechanisms. Similar illegal markets for CFCs were detected in Taiwan, Korea, and Hong Kong.

The Montreal Protocol is also expected to have effects on human health. A 2015 report by the U. S. Environmental Protection Agency estimates that the protection of the ozone layer under the treaty will prevent over 280 million cases of skin cancer, 1.5 million skin cancer deaths, and 45 million cataracts in the United States.

However, the hydrochlorofluorocarbons, or HCFCs, and hydrofluorocarbons, or HFCs, are now thought to contribute to anthropogenic global warming. On a molecule-for-molecule basis, these compounds are up to 10,000 times more potent greenhouse gases than carbon dioxide. The Montreal Protocol currently calls for a complete phase-out of HCFCs by 2030, but does not place any restriction on HFCs. Since the CFCs themselves are equally powerful greenhouse gases, the mere substitution of HFCs for CFCs does not significantly increase the rate of anthropogenic climate change, but over time a steady increase in their use could increase the danger that human activity will change the climate.

Policy experts have advocated for increased efforts to link ozone protection efforts to climate protection efforts. Policy decisions in one arena affect the costs and effectiveness of environmental improvements in the other.

Regional detections of non-compliance

In 2018, scientists monitoring the atmosphere following the 2010 phaseout date have reported evidence of continuing industrial production of CFC-11, likely in eastern Asia, with detrimental global effects on the ozone layer.[48][49] A monitoring study detected fresh atmospheric releases of carbon tetrachloride from China's Shandong province, beginning sometime after 2012, and accounting for a large part of emissions exceeding global estimates under the Montreal Protocol.

25th Anniversary Celebrations

The year 2012 marked the 25th anniversary of the signing of the Montreal Protocol. Accordingly, the Montreal Protocol community organized a range of celebrations at the national, regional and international levels to publicize its considerable success to date and to consider the work ahead for the future. Among its accomplishments are: The Montreal Protocol was the first international treaty to address a global environmental regulatory challenge; the first to embrace the "precautionary principle" in its design for science-based policymaking; the first treaty where independent experts on atmospheric science, environmental impacts, chemical technology, and economics, reported directly to Parties, without edit or censorship, functioning under norms of professionalism, peer review, and respect; the first to provide for national differences in responsibility and financial capacity to respond by establishing a multilateral fund for technology transfer; the first MEA with stringent reporting, trade, and binding chemical phase-out obligations for both developed and developing countries; and, the first treaty with a financial mechanism managed democratically by an Executive Board with equal representation by developed and developing countries.

Within 25 years of signing, parties to the MP celebrate significant milestones. Significantly, the world has phased-out 98% of the Ozone-Depleting Substances (ODS) contained in nearly 100 hazardous chemicals worldwide; every country is in compliance with stringent obligations; and, the MP has achieved the status of the first global regime with universal ratification; even the newest member state, South Sudan, ratified in 2013. UNEP received accolades for achieving global consensus that "demonstrates the world’s commitment to ozone protection, and more broadly, to global environmental protection".

Introduction to the Adjustments

The Second, Fourth, Seventh, Ninth, Eleventh and Nineteenth Meetings of the Parties to the Montreal Protocol on Substances that Deplete the Ozone Layer decided, on the basis of assessments made pursuant to Article 6 of the Protocol, to adopt adjustments and reductions of production and consumption of the controlled substances in

Annexes A, B, C and E to the Protocol as follows (the text here shows the cumulative effect of all the adjustments):

Article 2A: CFCs

1. Each Party shall ensure that for the twelve-month period commencing on the first day of the seventh month following the date of entry into force of this Protocol, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex A does not exceed its

calculated level of consumption in 1986. By the end of the same period, each Party producing one or more of these substances shall ensure that its calculated level of production of the substances does not exceed its calculated level of production in 1986, except that such level may have increased by no more than ten per cent based on the 1986 level. Such increase shall be permitted only so as to satisfy the basic domestic needs of the Parties operating under Article 5 and for the purposes of industrial rationalization between Parties.

2. Each Party shall ensure that for the period from 1 July 1991 to 31 December 1992 its calculated levels of consumption and production of the controlled substances in Group I of Annex A do not exceed 150 per cent of its calculated levels of production and consumption of those substances in 1986; with effect from 1 January 1993, the twelve-month control period for these controlled substances shall run from 1 January to 31 December each year.

3. Each Party shall ensure that for the twelve-month period commencing on 1 January 1994, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex A does not exceed, annually, twenty-five per cent of its calculated level of consumption in 1986. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed, annually, twenty-five per cent of its calculated level of production in 1986. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1986.

4. Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex A does not exceed zero. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by a quantity equal to the annual average of its production of the controlled substances in Group I of Annex A for basic domestic needs for the period 1995 to 1997 inclusive. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

5. Each Party shall ensure that for the twelve-month period commencing on 1 January 2003 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed eighty per cent of the annual average of its production of those substances for basic domestic needs for the period 1995 to 1997 inclusive.

6. Each Party shall ensure that for the twelve-month period commencing on 1 January 2005 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed fifty per cent of the annual average of its production of those substances for basic domestic needs for the period 1995 to 1997 inclusive.

7. Each Party shall ensure that for the twelve-month period commencing on 1 January 2007 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed fifteen per cent of the annual average of its production of those substances for basic domestic needs for the period 1995 to 1997 inclusive.

8. Each Party shall ensure that for the twelve-month period commencing on 1 January 2010 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed zero.

9. For the purposes of calculating basic domestic needs under paragraphs 4 to 8 of this Article, the calculation of the annual average of production by a Party includes any production entitlements that it has transferred in accordance with paragraph 5 of Article 2, and excludes any production entitlements that it has acquired in accordance with paragraph 5 of Article 2.

Article 2B: Halons

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1992, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group II of Annex A does not exceed, annually, its calculated level of consumption in 1986. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed, annually, its calculated level of production in 1986. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1986.

2. Each Party shall ensure that for the twelve-month period commencing on 1 January 1994, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group II of Annex A does not exceed zero. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may, until 1 January 2002 exceed that limit by up to fifteen per cent of its calculated level, it may exceed that limit by a quantity equal to the annual average of its production of the controlled substances in Group II of Annex A for basic domestic needs for the period 1995 to 1997 inclusive. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

3. Each Party shall ensure that for the twelve-month period commencing on 1 January 2005 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group II of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed fifty per cent of the annual average of its production of those substances for basic domestic needs for the period 1995 to 1997 inclusive.

4. Each Party shall ensure that for the twelve-month period commencing on 1 January 2010 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group II of Annex A for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed zero.

Article 2C: Other fully halogenated CFCs

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1993, its calculated level of consumption of the controlled substances in Group I of Annex B does not exceed, annually, eighty per cent of its calculated level of consumption in 1989. Each Party producing one or more of these substances shall, for the same period, ensure that its calculated level of production of the substances does not exceed, annually, eighty per cent of its calculated level of production in 1989. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated

level of production may exceed that limit by up to ten per cent of its calculated level of production in 1989.

2. Each Party shall ensure that for the twelve-month period commencing on 1 January 1994, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex B does not exceed, annually, twenty-five per cent of its calculated level of consumption in 1989. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed, annually, twenty-five per cent of its calculated level of production in 1989. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1989.

3. Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex B does not exceed zero. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may, until 1 January 2003 exceed that limit by up to fifteen per cent of its calculated level of production in 1989; thereafter, it may exceed that limit by a quantity equal to eighty per cent of the annual average of its production of the controlled substances in Group I of Annex B for basic domestic needs for the period 1998 to 2000 inclusive. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

4. Each Party shall ensure that for the twelve-month period commencing on 1 January 2007 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex B for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed fifteen per cent of the annual average of its production of those substances for basic domestic needs for the period 1998 to 2000 inclusive.

5. Each Party shall ensure that for the twelve-month period commencing on 1 January 2010 and in each twelvemonth period thereafter, its calculated level of production of the controlled substances in Group I of Annex B for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed zero.

Article 2D: Carbon tetrachloride

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1995, its calculated level of consumption of the controlled substance in Group II of Annex B does not exceed, annually, fifteen per cent of its calculated level of consumption in 1989. Each Party producing the substance shall, for the same period, ensure that its calculated level of production of the substance does not exceed, annually, fifteen per cent of its calculated level of production in 1989. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1989.

2. Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Group II of Annex B does not exceed zero. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to fifteen per cent of its calculated level of production in 1989.

This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

Article 2E: 1,1,1-Trichloroethane (Methyl chloroform)

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1993, its calculated level of consumption of the controlled substance in Group III of Annex B does not exceed, annually, its calculated level of consumption in 1989. Each Party producing the substance shall, for the same period, ensure that its calculated level of production of the substance does not exceed, annually, its calculated level of production in 1989. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1989.

2. Each Party shall ensure that for the twelve-month period commencing on 1 January 1994, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Group III of Annex B does not exceed, annually, fifty per cent of its calculated level of consumption in 1989. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed, annually, fifty per cent of its calculated level of production in 1989. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1989.

3. Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Group III of Annex B does not exceed zero. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to fifteen per cent of its calculated level of production for 1989. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

Article 2F: Hydrochlorofluorocarbons

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, the sum of:

(a) Two point eight per cent of its calculated level of consumption in 1989 of the controlled substances in Group I of Annex A; and

(b) Its calculated level of consumption in 1989 of the controlled substances in Group I of Annex C.

2. Each Party producing one or more of these substances shall ensure that for the twelve-month period commencing on 1 January 2004, and in each twelve-month period thereafter, its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, the average of:

(a) The sum of its calculated level of consumption in 1989 of the controlled substances in Group I of Annex C and two point eight per cent of its calculated level of consumption in 1989 of the controlled substances in Group I of Annex A; and

(b) The sum of its calculated level of production in 1989 of the controlled substances in Group I of Annex C and two point eight per cent of its calculated level of production in 1989 of the controlled substances in Group I of Annex A.

However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to fifteen per cent of its calculated level of production of the controlled substances in Group I of Annex C as defined above.

3. Each Party shall ensure that for the twelve month period commencing on 1 January 2004, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, sixty-five per cent of the sum referred to in paragraph 1 of this Article.

4. Each Party shall ensure that for the twelve-month period commencing on 1 January 2010, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, twenty-five per cent of the sum referred to in paragraph 1 of this Article. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, twenty-five per cent of the calculated level referred to in paragraph 2 of this Article. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production of the controlled substances in Group I of Annex C as referred to in paragraph 2.

5. Each Party shall ensure that for the twelve-month period commencing on 1 January 2015, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, ten per cent of the sum referred to in paragraph 1 of this Article. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, ten per cent of the calculated level referred to in paragraph 2 of this Article. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production of the controlled substances in Group I of Annex C as referred to in paragraph 2.

6. Each Party shall ensure that for the twelve-month period commencing on 1 January 2020, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed zero. Each Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed zero. However:

(a) Each Party may exceed that limit on consumption by up to zero point five per cent of the sum referred to in paragraph 1 of this Article in any such twelve-month period ending before 1 January 2030, provided that such consumption shall be restricted to the servicing of refrigeration and air-conditioning equipment existing on 1 January 2020;

(b) Each Party may exceed that limit on production by up to zero point five per cent of the average referred to in paragraph 2 of this Article in any such twelve-month period ending before 1 January 2030, provided that such production shall be restricted to the servicing of refrigeration and air-conditioning equipment existing on 1 January 2020.

7. As of 1 January 1996, each Party shall endeavour to ensure that:

(a) The use of controlled substances in Group I of Annex C is limited to those applications where other more environmentally suitable alternative substances or technologies are not available;

(b) The use of controlled substances in Group I of Annex C is not outside the areas of application currently met by controlled substances in Annexes A, B and C, except in rare cases for the protection of human life or human health; and

(c) Controlled substances in Group I of Annex C are selected for use in a manner that minimizes ozone depletion, in addition to meeting other environmental, safety and economic considerations.

Article 2G: Hydrobromofluorocarbons

Each Party shall ensure that for the twelve-month period commencing on 1 January 1996, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group II of Annex C does not exceed zero. Each Party producing the substances shall, for the same periods, ensure that its calculated level of production of the substances does not exceed zero. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

Article 2H: Methyl bromide

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 1995, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Annex E does not exceed, annually, its calculated level of consumption in 1991. Each Party producing the substance shall, for the same period, ensure that its calculated level of production of the substance does not exceed, annually, its calculated level of production in 1991. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1991.

2. Each Party shall ensure that for the twelve-month period commencing on 1 January 1999, and in the twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Annex E does not exceed, annually, seventy-five per cent of its calculated level of consumption in 1991. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed, annually, seventy-five per cent of its calculated level of production in 1991. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1991.

3. Each Party shall ensure that for the twelve-month period commencing on 1 January 2001, and in the twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Annex E does not exceed, annually, fifty per cent of its calculated level of consumption in 1991. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed, annually, fifty per cent of its calculated level of production in 1991. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1991.

4. Each Party shall ensure that for the twelve-month period commencing on 1 January 2003, and in the twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Annex E does not exceed, annually, thirty per cent of its calculated level of consumption in 1991. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the

substance does not exceed, annually, thirty per cent of its calculated level of production in 1991. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may exceed that limit by up to ten per cent of its calculated level of production in 1991.

5. Each Party shall ensure that for the twelve-month period commencing on 1 January 2005, and in each twelvemonth period thereafter, its calculated level of consumption of the controlled substance in Annex E does not exceed zero. Each Party producing the substance shall, for the same periods, ensure that its calculated level of production of the substance does not exceed zero. However, in order to satisfy the basic domestic needs of the Parties operating under paragraph 1 of Article 5, its calculated level of production may, until 1 January 2002 exceed that limit by up to fifteen per cent of its calculated level of production in 1991; thereafter, it may exceed that limit by a quantity equal to the annual average of its production of the controlled substance in Annex E for basic domestic needs for the period 1995 to 1998 inclusive. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be critical uses.

5 bis. Each Party shall ensure that for the twelve-month period commencing on 1 January 2005 and in each twelvemonth period thereafter, its calculated level of production of the controlled substance in Annex E for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed eighty per cent of the annual average of its production of the substance for basic domestic needs for the period 1995 to 1998 inclusive.

5 ter. Each Party shall ensure that for the twelve-month period commencing on 1 January 2015 and in each twelvemonth period thereafter, its calculated level of production of the controlled substance in Annex E for the basic domestic needs of the Parties operating under paragraph 1 of Article 5 does not exceed zero.

6. The calculated levels of consumption and production under this Article shall not include the amounts used by the Party for quarantine and pre-shipment applications.

Article 2I: Bromochloromethane

Each Party shall ensure that for the twelve-month period commencing on 1 January 2002, and in each twelve-month period thereafter, its calculated level of consumption and production of the controlled substance in Group III of Annex C does not exceed zero. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be essential.

Article 2J: Hydrofluorocarbons

1. Each Party shall ensure that for the twelve-month period commencing on 1 January 2019, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Annex F, expressed in CO2 equivalents, does not exceed the percentage, set out for the respective range of years specified in subparagraphs (a) to (e) below, of the annual average of its calculated levels of consumption of Annex F controlled substances for the years 2011, 2012 and 2013, plus fifteen per cent of its calculated level of consumption of Annex C, Group I, controlled substances as set out in paragraph 1 of Article 2F, expressed in CO2 equivalents:

(a) 2019 to 2023: 90 per cent

(b) 2024 to 2028: 60 per cent

(c) 2029 to 2033: 30 per cent

(d) 2034 to 2035: 20 per cent

(e) 2036 and thereafter: 15 per cent

2. Notwithstanding paragraph 1 of this Article, the Parties may decide that a Party shall ensure that, for the twelve-month period commencing on 1 January 2020, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Annex F, expressed in CO_2 equivalents, does not exceed the percentage, set out for the respective range of years specified in subparagraphs (a) to (e) below, of the annual average of its calculated levels of consumption of Annex F controlled substances for the years 2011, 2012 and 2013, plus twenty-five per cent of its calculated level of consumption of Annex C, Group I, controlled substances as set out in paragraph 1 of Article 2F, expressed in CO_2 equivalents:

(a) 2020 to 2024: 95 per cent

(b) 2025 to 2028: 65 per cent

(c) 2029 to 2033: 30 per cent

(d) 2034 to 2035: 20 per cent

(e) 2036 and thereafter: 15 per cent

3. Each Party producing the controlled substances in Annex F shall ensure that for the twelve-month period commencing on 1 January 2019, and in each twelve-month period thereafter, its calculated level of production of the controlled substances in Annex F, expressed in CO_2 equivalents, does not exceed the percentage, set out for the respective range of years specified in subparagraphs (a) to (e) below, of the annual average of its calculated levels of production of Annex F controlled substances for the years 2011, 2012 and 2013, plus fifteen per cent of its calculated level of production of Annex C, Group I, controlled substances as set out in paragraph 2 of Article 2F, expressed in CO_2 equivalents:

(a) 2019 to 2023: 90 per cent

(b) 2024 to 2028: 60 per cent

(c) 2029 to 2033: 30 per cent

(d) 2034 to 2035: 20 per cent

(e) 2036 and thereafter: 15 per cent

4. Notwithstanding paragraph 3 of this Article, the Parties may decide that a Party producing the controlled substances in Annex F shall ensure that for the twelve-month period commencing on 1 January 2020, and in each twelve-month period thereafter, its calculated level of production of the controlled substances in Annex F, expressed in CO_2 equivalents, does not exceed the percentage, set out for the respective range of years specified in subparagraphs (a) to (e) below, of the annual average of its calculated levels of production of Annex F controlled substances for the years 2011, 2012 and 2013, plus twenty-five per cent of its calculated level of production of Annex C, Group I, controlled substances as set out in paragraph 2 of Article 2F, expressed in CO_2 equivalents:

(a) 2020 to 2024: 95 per cent

(b) 2025 to 2028: 65 per cent

(c) 2029 to 2033: 30 per cent

(d) 2034 to 2035: 20 per cent

(e) 2036 and thereafter: 15 per cent

5. Paragraphs 1 to 4 of this Article will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by the Parties to be exempted uses.

6. Each Party manufacturing Annex C, Group I, or Annex F substances shall ensure that for the twelve-month period commencing on 1 January 2020, and in each twelve-month period thereafter, its emissions of Annex F, Group II, substances generated in each production facility that manufactures Annex C, Group I, or Annex F substances are destroyed to the extent practicable using technology approved by the Parties in the same twelve-month period.

7. Each Party shall ensure that any destruction of Annex F, Group II, substances generated by facilities that produce Annex C, Group I, or Annex F substances shall occur only by technologies approved by the Parties.

Article 3: Calculation of control levels

1. For the purposes of Articles 2, 2A to 2J and 5, each Party shall, for each group of substances in Annex A, Annex B, Annex C, Annex E or Annex F, determine its calculated levels of: For the purposes of Articles 2, 2A to 2I and 5, each Party shall, for each group of substances in Annex A, Annex B, Annex C or Annex E determine its calculated levels of:

(a) Production by:

(i) multiplying its annual production of each controlled substance by the ozone depleting potential specified in respect of it in Annex A, Annex B, Annex C or Annex E;, except as otherwise specified in paragraph 2;

(ii) adding together, for each such Group, the resulting figures;

(b) Imports and exports, respectively, by following, mutatis mutandis, the procedure set out in subparagraph (a); and

(c) Consumption by adding together its calculated levels of production and imports and subtracting its calculated level of exports as determined in accordance with subparagraphs (a) and (b). However, beginning on 1 January 1993, any export of controlled substances to non-Parties shall not be subtracted in calculating the consumption level of the exporting Party; and (d) Emissions of Annex F, Group II, substances generated in each facility that generates Annex C, Group I, or Annex F substances by including, among other things, amounts emitted from equipment leaks, process vents and destruction devices, but excluding amounts captured for use, destruction or storage.

2. When calculating levels, expressed in CO2 equivalents, of production, consumption, imports, exports and emissions of Annex F and Annex C, Group I, substances for the purposes of Article 2J, paragraph 5

bis of Article 2 and paragraph 1 (d) of Article 3, each Party shall use the global warming potentials of those substances specified in Group I of Annex A, Annex C and Annex F.

Article 4: Control of trade with non-Parties

1. As of 1 January 1990, each party shall ban the import of the controlled substances in Annex A from any State not party to this Protocol. 1 bis. Within one year of the date of the entry into force of this paragraph, each Party shall ban the import of the controlled substances in Annex B from any State not party to this Protocol.

1 ter. Within one year of the date of entry into force of this paragraph, each Party shall ban the import of any controlled substances in Group II of Annex C from any State not party to this Protocol.

1 qua. Within one year of the date of entry into force of this paragraph, each Party shall ban the import of the controlled substance in Annex E from any State not party to this Protocol.

1 quin.As of 1 January 2004, each Party shall ban the import of the controlled substances in Group I of Annex C from any State not party to this Protocol.

1 sex. Within one year of the date of entry into force of this paragraph, each Party shall ban the import of the controlled substance in Group III of Annex C from any State not party to this Protocol.

1 sept. Upon entry into force of this paragraph, each Party shall ban the import of the controlled substances in Annex F from any State not Party to this Protocol.

2. As of 1 January 1993, each Party shall ban the export of any controlled substances in Annex A to any State not party to this Protocol.

2 bis. Commencing one year after the date of entry into force of this paragraph, each Party shall ban the export of any controlled substances in Annex B to any State not party to this Protocol.

2 ter. Commencing one year after the date of entry into force of this paragraph, each Party shall ban the export of any controlled substances in Group II of Annex C to any State not party to this Protocol.

2 qua. Commencing one year of the date of entry into force of this paragraph, each Party shall ban the export of the

controlled substance in Annex E to any State not party to this Protocol.

2 quin.As of 1 January 2004, each Party shall ban the export of the controlled substances in Group I of Annex C to any State not party to this Protocol.

2 sex. Within one year of the date of entry into force of this paragraph, each Party shall ban the export of the controlled substance in Group III of Annex C to any State not party to this Protocol.

2 sept. Upon entry into force of this paragraph, each Party shall ban the export of the controlled substances in Annex F to any State not Party to this Protocol.

3. By 1 January 1992, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of products containing controlled substances in Annex A. Parties that have not objected to the annex in accordance with those procedures shall ban, within one year of the annex having become effective,the import of those products from any State not party to this Protocol.

3 bis. Within three years of the date of the entry into force of this paragraph, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of products containing controlled substances in Annex B. Parties that have not objected to the annex in accordance with those procedures shall ban, within one year of the annex having become effective, the import of those products from any State not party to this Protocol.

3 ter. Within three years of the date of entry into force of this paragraph, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of products containing controlled substances in Group II of Annex C. Parties that have not objected to the annex in accordance with those procedures shall ban, within one year of the annex having become effective, the import of those products from any State not party to this Protocol.

4. By 1 January 1994, the Parties shall determine the feasibility of banning or restricting, from States not party to this Protocol, the import of products produced with, but not containing, controlled substances in Annex A.

If determined feasible, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of such products. Parties that have not objected to the annex in accordance with those procedures shall ban or restrict, within one year of the annex having become effective, the import of those products from any State not party to this Protocol.

4 bis. Within five years of the date of the entry into force of this paragraph, the Parties shall determine the feasibility of banning or restricting, from States not party to this Protocol, the import of products produced with, but not containing, controlled substances in Annex B. If determined feasible, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of such products.

Parties that have not objected to the annex in accordance with those procedures shall ban or restrict, within one year of the annex having become effective, the import of those products from any State not party to this Protocol.

4 ter. Within five years of the date of entry into force of this paragraph, the Parties shall determine the feasibility of banning or restricting, from States not party to this Protocol, the import of products produced with, but not containing, controlled substances in Group II of Annex C. If determined feasible, the Parties shall, following the procedures in Article 10 of the Convention, elaborate in an annex a list of such products. Parties that have not objected to the annex in accordance with those procedures shall ban or restrict, within one year of the annex having become effective, the import of those products from any State not party to this Protocol.

5. Each Party undertakes to the fullest practicable extent to discourage the export to any State not party to this Protocol of technology for producing and for utilizing controlled substances in Annexes A, B, C, E and F Annexes A, B, C and E.

6. Each Party shall refrain from providing new subsidies, aid, credits, guarantees or insurance programmes for the export to States not party to this Protocol of products, equipment, plants or technology that would facilitate the production of controlled substances in Annexes A, B, C, E and F Annexes A, B, C and E.

7. Paragraphs 5 and 6 shall not apply to products, equipment, plants or technology that improve the containment, recovery, recycling or destruction of controlled substances, promote the development of

alternative substances, or otherwise contribute to the reduction of emissions of controlled substances in Annexes A, B, C, E and F Annexes A, B, C and E.

8. Notwithstanding the provisions of this Article, imports and exports referred to in paragraphs 1 to 4 ter of this Article may be permitted from, or to, any State not party to this Protocol, if that State is determined, by a meeting of the Parties, to be in full compliance with Article 2, Articles 2A to 2J Articles 2A to 2I and this Article, and have submitted data to that effect as specified in Article 7.

9. For the purposes of this Article, the term "State not party to this Protocol" shall include, with respect to a particular controlled substance, a State or regional economic integration organization that has not agreed to be bound by the control measures in effect for that substance.

10. By 1 January 1996, the Parties shall consider whether to amend this Protocol in order to extend the measures in this Article to trade in controlled substances in Group I of Annex C and in Annex E with States not party to the Protocol.

Article 4A: Control of trade with Parties

1. Where, after the phase-out date applicable to it for a controlled substance, a Party is unable, despite having

taken all practicable steps to comply with its obligation under the Protocol, to cease production of that

substance for domestic consumption, other than for uses agreed by the Parties to be essential, it shall ban the

export of used, recycled and reclaimed quantities of that substance, other than for the purpose of destruction.

2. Paragraph 1 of this Article shall apply without prejudice to the operation of Article 11 of the Convention and

the non-compliance procedure developed under Article 8 of the Protocol.

Article 4B: Licensing

1. Each Party shall, by 1 January 2000 or within three months of the date of entry into force of this Article for it, whichever is the later, establish and implement a system for licensing the import and export of new, used, recycled and reclaimed controlled substances in Annexes A, B, C and E.

2. Notwithstanding paragraph 1 of this Article, any Party operating under paragraph 1 of Article 5 which decides it is not in a position to establish and implement a system for licensing the import and export of controlled substances in Annexes C and E, may delay taking those actions until 1 January 2005 and 1 January 2002, respectively.

2 bis. Each Party shall, by 1 January 2019 or within three months of the date of entry into force of this paragraph for it, whichever is later, establish and implement a system for licensing the import and export of new, used, recycled and reclaimed controlled substances in Annex F. Any Party operating under paragraph 1 of Article 5 that decides it is not in a position to establish and implement such a system by 1 January 2019 may delay taking those actions until 1 January 2021.

3. Each Party shall, within three months of the date of introducing its licensing system, report to the Secretariat

on the establishment and operation of that system.

4. The Secretariat shall periodically prepare and circulate to all Parties a list of the Parties that have reported to

it on their licensing systems and shall forward this information to the Implementation Committee for

consideration and appropriate recommendations to the Parties.

Article 5: Special situation of developing countries

1. Any Party that is a developing country and whose annual calculated level of consumption of the controlled

substances in Annex A is less than 0.3 kilograms per capita on the date of the entry into force of the Protocol

for it, or any time thereafter until 1 January 1999, shall, in order to meet its basic domestic needs, be entitled

to delay for ten years its compliance with the control measures set out in Articles 2A to 2E, provided that any

further amendments to the adjustments or Amendment adopted at the Second Meeting of the Parties in

London, 29 June 1990, shall apply to the Parties operating under this paragraph after the review provided for

in paragraph 8 of this Article has taken place and shall be based on the conclusions of that review.

1 bis. The Parties shall, taking into account the review referred to in paragraph 8 of this Article, the assessments

made pursuant to Article 6 and any other relevant information, decide by 1 January 1996, through the

procedure set forth in paragraph 9 of Article 2:

(a) With respect to paragraphs 1 to 6 of Article 2F, what base year, initial levels, control schedules and

phase-out date for consumption of the controlled substances in Group I of Annex C will apply to Parties

operating under paragraph 1 of this Article;

(b) With respect to Article 2G, what phase-out date for production and consumption of the controlled

substances in Group II of Annex C will apply to Parties operating under paragraph 1 of this Article; and

(c) With respect to Article 2H, what base year, initial levels and control schedules for consumption and production of the controlled substance in Annex E will apply to Parties operating under paragraph 1 of this Article.

2. However, any Party operating under paragraph 1 of this Article shall exceed neither an annual calculated level of consumption of the controlled substances in Annex A of 0.3 kilograms per capita nor an annual calculated level of consumption of controlled substances of Annex B of 0.2 kilograms per capita.

3. When implementing the control measures set out in Articles 2A to 2E, any Party operating under paragraph 1 of this Article shall be entitled to use:

(a) For controlled substances under Annex A, either the average of its annual calculated level of consumption for the period 1995 to 1997 inclusive or a calculated level of consumption of 0.3 kilograms per capita, whichever is the lower, as the basis for determining its compliance with the control measures relating to consumption.

(b) For controlled substances under Annex B, the average of its annual calculated level of consumption for the period 1998 to 2000 inclusive or a calculated level of consumption of 0.2 kilograms per capita, whichever is the lower, as the basis for determining its compliance with the control measures relating to consumption.

(c) For controlled substances under Annex A, either the average of its annual calculated level of production for the period 1995 to 1997 inclusive or a calculated level of production of 0.3 kilograms per capita, whichever is the lower, as the basis for determining its compliance with the control measures relating to production.

(d) For controlled substances under Annex B, either the average of its annual calculated level of production for the period 1998 to 2000 inclusive or a calculated level of production of 0.2 kilograms per capita, whichever is the lower, as the basis for determining its compliance with the control measures relating to production.

4. If a Party operating under paragraph 1 of this Article, at any time before the control measures obligations in

Articles 2A to 2I 2J become applicable to it, finds itself unable to obtain an adequate supply of controlled

substances, it may notify this to the Secretariat. The Secretariat shall forthwith transmit a copy of such

notification to the Parties, which shall consider the matter at their next Meeting, and decide upon appropriate

action to be taken.

5. Developing the capacity to fulfil the obligations of the Parties operating under paragraph 1 of this Article to

comply with the control measures set out in Articles 2A to 2E and Articles 2I and 2J Article 2I, and with

any control measures in Articles 2F to 2H that are decided pursuant to paragraph 1 bis of this Article, and

their implementation by those same Parties will depend upon the effective implementation of the financial

co-operation as provided by Article 10 and the transfer of technology as provided by Article 10A.

6. Any Party operating under paragraph 1 of this Article may, at any time, notify the Secretariat in writing that,

having taken all practicable steps it is unable to implement any or all of the obligations laid down in Articles

2A to 2E and Articles 2I and 2J Article 2I, or any or all obligations in Articles 2F to 2H that are decided

pursuant to paragraph 1 bis of this Article, due to the inadequate implementation of Articles 10 and 10A. The

Secretariat shall forthwith transmit a copy of the notification to the Parties, which shall consider the matter at

their next Meeting, giving due recognition to paragraph 5 of this Article and shall decide upon appropriate

action to be taken.

7. During the period between notification and the Meeting of the Parties at which the appropriate action referred

to in paragraph 6 above is to be decided, or for a further period if the Meeting of the Parties so decides, the

non-compliance procedures referred to in Article 8 shall not be invoked against the notifying Party.

8. A Meeting of the Parties shall review, not later than 1995, the situation of the Parties operating under

paragraph 1 of this Article, including the effective implementation of financial co-operation and transfer of

technology to them, and adopt such revisions that may be deemed necessary regarding the schedule of control

measures applicable to those Parties.

8 bis. Based on the conclusions of the review referred to in paragraph 8 above:

(a) With respect to the controlled substances in Annex A, a Party operating under paragraph 1 of this Article shall, in order to meet its basic domestic needs, be entitled to delay for ten years its compliance with the control measures adopted by the Second Meeting of the Parties in London, 29 June 1990, and reference by the Protocol to Articles 2A and 2B shall be read accordingly;

(b) With respect to the controlled substances in Annex B, a Party operating under paragraph 1 of this Article shall, in order to meet its basic domestic needs, be entitled to delay for ten years its compliance with the control measures adopted by the Second Meeting of the Parties in London, 29 June 1990, and reference by the Protocol to Articles 2C to 2E shall be read accordingly.

8 ter. Pursuant to paragraph 1 bis above:

(a) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2013, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, the average of its calculated levels of consumption in 2009 and 2010. Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2013 and in each twelve-month period thereafter, its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, the average of its calculated levels of production in 2009 and 2010;

(b) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2015, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, ninety per cent of the average of its calculated levels of consumption in 2009 and 2010. Each such Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, ninety per cent of the average of its calculated levels of production in 2009 and 2010;

(c) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2020, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, sixty-five per cent of the average of its calculated levels of consumption in 2009 and 2010. Each such Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, sixty-five per cent of the average of its calculated levels of production in 2009 and 2010;

(d) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2025, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed, annually, thirty-two point five per cent of the average of its calculated levels of consumption in 2009 and 2010. Each such Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed, annually, thirty-two point five per cent of the average of its calculated levels of production in 2009 and 2010;

(e) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2030, and in each twelve-month period thereafter, its calculated level of consumption of the controlled substances in Group I of Annex C does not exceed zero. Each such Party producing one or more of these substances shall, for the same periods, ensure that its calculated level of production of the controlled substances in Group I of Annex C does not exceed zero. However:

(i) Each such Party may exceed that limit on consumption in any such twelve-month period so long as the sum of its calculated levels of consumption over the ten-year period from 1 January 2030 to 1 January 2040, divided by ten, does not exceed two point five per cent of the average of its calculated levels of consumption in 2009 and 2010, and provided that such consumption shall be restricted to the servicing of refrigeration and air-conditioning equipment existing on 1 January 2030;

(ii) Each such Party may exceed that limit on production in any such twelve-month period so long as the sum of its calculated levels of production over the ten-year period from 1 January 2030 to 1 January 2040, divided by ten, does not exceed two point five per cent of the average of its calculated levels of production in 2009 and 2010, and provided that such production shall be restricted to the servicing of refrigeration and air-conditioning equipment existing on 1 January 2030.

(f) Each Party operating under paragraph 1 of this Article shall comply with Article 2G;

(g) With regard to the controlled substance contained in Annex E:

(i) As of 1 January 2002 each Party operating under paragraph 1 of this Article shall comply with the control measures set out in paragraph 1 of Article 2H and, as the basis for its compliance with these control measures, it shall use the average of its annual calculated level of consumption and production, respectively, for the period of 1995 to 1998 inclusive;

(ii) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2005, and in each twelve-month period thereafter, its calculated levels of consumption and production of the controlled substance in Annex E do not exceed, annually, eighty per cent of the average of its annual calculated levels of consumption and production, respectively, for the period of 1995 to 1998 inclusive;

(iii) Each Party operating under paragraph 1 of this Article shall ensure that for the twelve-month period commencing on 1 January 2015 and in each twelve-month period thereafter, its calculated levels of consumption and production of the controlled substance in Annex E do not exceed zero. This paragraph will apply save to the extent that the Parties decide to permit the level of production or consumption that is necessary to satisfy uses agreed by them to be critical uses;

(iv) The calculated levels of consumption and production under this subparagraph shall not include the amounts used by the Party for quarantine and pre-shipment applications.

8 qua

(a) Each Party operating under paragraph 1 of this Article, subject to any adjustments made to the control measures in Article 2J in accordance with paragraph 9 of Article 2, shall be entitled to delay its compliance with the control measures set out in subparagraphs (a) to (e) of paragraph 1 of Article 2J and subparagraphs (a) to (e) of paragraph 3 of Article 2J and modify those measures as follows:

(i) 2024 to 2028: 100 per cent

(ii) 2029 to 2034: 90 per cent

(iii) 2035 to 2039: 70 per cent

(iv) 2040 to 2044: 50 per cent

(v) 2045 and thereafter: 20 per cent

(b) Notwithstanding subparagraph (a) above, the Parties may decide that a Party operating under paragraph 1 of this Article, subject to any adjustments made to the control measures in Article 2J in accordance with paragraph 9 of Article 2, shall be entitled to delay its compliance with the control measures set out in subparagraphs (a) to (e) of paragraph 1 of Article 2J and subparagraphs (a) to (e) of paragraph 3 of Article 2J and modify those measures as follows:

(i) 2028 to 2031: 100 per cent

(ii) 2032 to 2036: 90 per cent

(iii) 2037 to 2041: 80 per cent

(iv) 2042 to 2046: 70 per cent

(v) 2047 and thereafter: 15 per cent

(c) Each Party operating under paragraph 1 of this Article, for the purposes of calculating its consumption baseline under Article 2J, shall be entitled to use the average of its calculated levels of consumption of Annex F controlled substances for the years 2020, 2021 and 2022, plus sixtyfive per cent of its baseline consumption of Annex C, Group I, controlled substances as set out in paragraph 8 ter of this Article.

(d) Notwithstanding subparagraph (c) above, the Parties may decide that a Party operating under paragraph 1 of this Article, for the purposes of calculating its consumption baseline under Article 2J, shall be entitled to use the average of its calculated levels of consumption of Annex F controlled substances for the years 2024, 2025 and 2026, plus sixty-five per cent of its baseline consumption of Annex C, Group I, controlled substances as set out in paragraph 8 ter of this Article.

(e) Each Party operating under paragraph 1 of this Article and producing the controlled substances in Annex F, for the purposes of calculating its production baseline under Article 2J, shall be entitled to use the average of its calculated levels of production of Annex F controlled substances for the years 2020, 2021 and 2022, plus sixty-five per cent of its baseline production of Annex C, Group I, controlled substances as set out in paragraph 8 ter of this Article.

(f) Notwithstanding subparagraph (e) above, the Parties may decide that a Party operating under paragraph 1 of this Article and producing the controlled substances in Annex F, for the purposes of calculating its production baseline under Article 2J, shall be entitled to use the average of its calculated levels of production of Annex F controlled substances for the years 2024, 2025 and 2026, plus sixty-five per cent of its baseline production of Annex C, Group I, controlled substances as set out in paragraph 8 ter of this Article.

(g) Subparagraphs (a) to (f) of this paragraph will apply to calculated levels of production and consumption save to the extent that a high-ambient-temperature exemption applies based on criteria decided by the Parties.

9**. Decisions of the Parties referred to in paragraph 4, 6 and 7 of this Article shall be taken according to the same**

procedure applied to decision-making under Article 10.

Article 6: Assessment and review of control measures

Beginning in 1990, and at least every four years thereafter, the Parties shall assess the control measures provided for in Article 2 and Articles 2A to 2J Articles 2A to 2I on the basis of available scientific, environmental, technical and economic information. At least one year before each assessment, the Parties shall convene appropriate panels of experts qualified in the fields mentioned and determine the composition and terms of reference of any such panels. Within one year of being convened, the panels will report their conclusions, through the Secretariat, to the Parties.

Article 7: Reporting of data

1. Each Party shall provide to the Secretariat, within three months of becoming a Party, statistical data on its production, imports and exports of each of the controlled substances in Annex A for the year 1986, or the best possible estimates of such data where actual data are not available.

2. Each Party shall provide to the Secretariat statistical data on its production, imports and exports of each of the controlled substances

– in Annex B and Groups I and II of Annex C for the year 1989;

– in Annex E, for the year 1991,

– in Annex F, for the years 2011 to 2013, except that Parties operating under paragraph 1 of Article 5 shall provide such data for the years 2020 to 2022, but those Parties operating under paragraph 1 of Article 5 to which subparagraphs (d) and (f) of paragraph 8 qua of Article 5 applies shall provide such data for the years 2024 to 2026;

or the best possible estimates of such data where actual data are not available, not later than three months after the date when the provisions set out in the Protocol with regard to the substances in Annexes B, C, E and F C and E respectively enter into force for that Party.

3. Each Party shall provide to the Secretariat statistical data on its annual production (as defined in paragraph 5 of Article 1) of each of the controlled substances listed in Annexes A, B, C, E and F C and E and, separately, for each substance,

– Amounts used for feedstocks,

– Amounts destroyed by technologies approved by the Parties, and

– Imports from and exports to Parties and non-Parties respectively,

for the year during which provisions concerning the substances in Annexes A, B, C, E and F C and E

respectively entered into force for that Party and for each year thereafter. Each Party shall provide to the Secretariat statistical data on the annual amount of the controlled substance listed in Annex E used for quarantine and pre-shipment applications. Data shall be forwarded not later than nine months after the end of the year to which the data relate.

3 bis. Each Party shall provide to the Secretariat separate statistical data of its annual imports and exports of each of the controlled substances listed in Group II of Annex A and Group I of Annex C that have been recycled.

3 ter. Each Party shall provide to the Secretariat statistical data on its annual emissions of Annex F, Group II, controlled substances per facility in accordance with paragraph 1 (d) of Article 3 of the Protocol.

4. For Parties operating under the provisions of paragraph 8 (a) of Article 2, the requirements in paragraphs 1, 2,

3 and 3 bis of this Article in respect of statistical data on production, imports and exports shall be satisfied if the regional economic integration organization concerned provides data on production, imports and exports between the organization and States that are not members of that organization.

Article 8: Non-compliance

The Parties, at their first meeting, shall consider and approve procedures and institutional mechanisms for determining non-compliance with the provisions of this Protocol and for treatment of Parties found to be in noncompliance.

Article 9: Research, development, public awareness and exchange of information

1. The Parties shall co-operate, consistent with their national laws, regulations and practices and taking into account in particular the needs of developing countries, in promoting, directly or through competent international bodies, research, development and exchange of information on:

(a) best technologies for improving the containment, recovery, recycling, or destruction of controlled substances or otherwise reducing their emissions;

(b) possible alternatives to controlled substances, to products containing such substances, and to products manufactured with them; and

(c) costs and benefits of relevant control strategies.

2. The Parties, individually, jointly or through competent international bodies, shall co-operate in promoting public awareness of the environmental effects of the emissions of controlled substances and other substances that deplete the ozone layer.

3. Within two years of the entry into force of this Protocol and every two years thereafter, each Party shall submit to the Secretariat a summary of the activities it has conducted pursuant to this Article.

Article 10: Financial mechanism

1. The Parties shall establish a mechanism for the purposes of providing financial and technical co-operation, including the transfer of technologies, to Parties operating under paragraph 1 of Article 5 of this Protocol to enable their compliance with the control measures set out in Articles 2A to 2E, Article 2I and Article 2J and Article 2I, and any control measures in Articles 2F to 2H that are decided pursuant to paragraph 1 bis of Article 5 of the Protocol. The mechanism, contributions to which shall be additional to other financial transfers to Parties operating under that paragraph, shall meet all agreed incremental costs of such Parties in order to enable their compliance with the control measures of the Protocol. An indicative list of the categories of incremental costs shall be decided by the meeting of the Parties. Where a Party operating under paragraph 1 of Article 5 chooses to avail itself of funding from any other financial mechanism that could result in meeting any part of its agreed incremental costs, that part shall not be met by the financial mechanism under Article 10 of this Protocol.

2. The mechanism established under paragraph 1 shall include a Multilateral Fund. It may also include other means of multilateral, regional and bilateral co-operation.

3. The Multilateral Fund shall:

(a) Meet, on a grant or concessional basis as appropriate, and according to criteria to be decided upon by the Parties, the agreed incremental costs;

(b) Finance clearing-house functions to:

(i) Assist Parties operating under paragraph 1 of Article 5, through country specific studies and other technical co-operation, to identify their needs for co-operation;

(ii) Facilitate technical co-operation to meet these identified needs;

(iii) Distribute, as provided for in Article 9, information and relevant materials, and hold workshops, training sessions, and other related activities, for the benefit of Parties that are developing countries; and

(iv) Facilitate and monitor other multilateral, regional and bilateral co-operation available to Parties that are developing countries;

(c) Finance the secretarial services of the Multilateral Fund and related support costs.

4. The Multilateral Fund shall operate under the authority of the Parties who shall decide on its overall policies.

5. The Parties shall establish an Executive Committee to develop and monitor the implementation of specific operational policies, guidelines and administrative arrangements, including the disbursement of resources, for the purpose of achieving the objectives of the Multilateral Fund. The Executive Committee shall discharge its tasks and responsibilities, specified in its terms of reference as agreed by the Parties, with the co-operation and assistance of the International Bank for Reconstruction and Development (World Bank), the United Nations Environment Programme, the United Nations Development Programme or other appropriate agencies depending on their respective areas of expertise. The members of the Executive Committee, which shall be selected on the basis of a balanced representation of the Parties operating under paragraph 1 of

Article 5 and of the Parties not so operating, shall be endorsed by the Parties.

6. The Multilateral Fund shall be financed by contributions from Parties not operating under paragraph 1 of Article 5 in convertible currency or, in certain circumstances, in kind and/or in national currency, on the basis of the United Nations scale of assessments. Contributions by other Parties shall be encouraged. Bilateral and, in particular cases agreed by a decision of the Parties, regional co-operation may, up to a percentage and consistent with any criteria to be specified by decision of the Parties, be considered as a contribution to the Multilateral Fund, provided that such co-operation, as a minimum:

(a) Strictly relates to compliance with the provisions of this Protocol;

(b) Provides additional resources; and

(c) Meets agreed incremental costs.

7. The Parties shall decide upon the programme budget of the Multilateral Fund for each fiscal period and upon the percentage of contributions of the individual Parties thereto.

8. Resources under the Multilateral Fund shall be disbursed with the concurrence of the beneficiary Party.

9. Decisions by the Parties under this Article shall be taken by consensus whenever possible. If all efforts at consensus have been exhausted and no agreement reached, decisions shall be adopted by a two-thirds majority vote of the Parties present and voting, representing a majority of the Parties operating under paragraph 1 of Article 5 present and voting and a majority of the Parties not so operating present and

voting. 10. The financial mechanism set out in this Article is without prejudice to any future arrangements that may be developed with respect to other environmental issues.

Article 10A: Transfer of technology

Each Party shall take every practicable step, consistent with the programmes supported by the financial mechanism, to ensure:

(a) that the best available, environmentally safe substitutes and related technologies are expeditiously transferred to Parties operating under paragraph 1 of Article 5; and

(b) that the transfers referred to in subparagraph (a) occur under fair and most favourable conditions.

Article 11: Meetings of the parties

1. The Parties shall hold meetings at regular intervals. The Secretariat shall convene the first meeting of the Parties not later than one year after the date of the entry into force of this Protocol and in conjunction with a meeting of the Conference of the Parties to the Convention, if a meeting of the latter is scheduled within that period.

2. Subsequent ordinary meetings of the parties shall be held, unless the Parties otherwise decide, in conjunction with meetings of the Conference of the Parties to the Convention. Extraordinary meetings of the Parties shall be held at such other times as may be deemed necessary by a meeting of the Parties, or at the written request of any Party, provided that within six months of such a request being communicated to them by the Secretariat, it is supported by at least one third of the Parties.

3. The Parties, at their first meeting, shall:

(a) adopt by consensus rules of procedure for their meetings;

(b) adopt by consensus the financial rules referred to in paragraph 2 of Article 13;

(c) establish the panels and determine the terms of reference referred to in Article 6;

(d) consider and approve the procedures and institutional mechanisms specified in Article 8; and

(e) begin preparation of workplans pursuant to paragraph 3 of Article 10.

[The Article 10 in question is that of the original Protocol adopted in 1987.]

4. The functions of the meetings of the Parties shall be to:

(a) review the implementation of this Protocol;

(b) decide on any adjustments or reductions referred to in paragraph 9 of Article 2;

(c) decide on any addition to, insertion in or removal from any annex of substances and on related control

measures in accordance with paragraph 10 of Article 2;

(d) establish, where necessary, guidelines or procedures for reporting of information as provided for in Article 7 and paragraph 3 of Article 9;

(e) review requests for technical assistance submitted pursuant to paragraph 2 of Article 10;

(f) review reports prepared by the secretariat pursuant to subparagraph (c) of Article 12;

(g) assess, in accordance with Article 6, the control measures;

(h) consider and adopt, as required, proposals for amendment of this Protocol or any annex and for any ew annex;

(i) consider and adopt the budget for implementing this Protocol; and

(j) consider and undertake any additional action that may be required for the achievement of the purposes of this Protocol.

5. The United Nations, its specialized agencies and the International Atomic Energy Agency, as well as any State not party to this Protocol, may be represented at meetings of the Parties as observers. Anybody or agency, whether national or international, governmental or non-governmental, qualified in fields relating to the protection of the ozone layer which has informed the secretariat of its wish to be represented at a meeting of the Parties as an observer may be admitted unless at least one third of the Parties present object. The admission and participation of observers shall be subject to the rules of procedure adopted by the Parties.

Article 12: Secretariat

For the purposes of this Protocol, the Secretariat shall:

(a) arrange for and service meetings of the Parties as provided for in Article 11;

(b) receive and make available, upon request by a Party, data provided pursuant to Article 7;

(c) prepare and distribute regularly to the Parties reports based on information received pursuant to Articles 7 and 9;

(d) notify the Parties of any request for technical assistance received pursuant to Article 10 so as to facilitate the provision of such assistance;

(e) encourage non-Parties to attend the meetings of the Parties as observers and to act in accordance with the provisions of this Protocol;

(f) provide, as appropriate, the information and requests referred to in subparagraphs (c) and (d) to such non-party observers; and

(g) perform such other functions for the achievement of the purposes of this Protocol as may be assigned to it by the Parties.

Article 13: Financial provisions

1. The funds required for the operation of this Protocol, including those for the functioning of the Secretariat related to this Protocol, shall be charged exclusively against contributions from the Parties.

2. The Parties, at their first meeting, shall adopt by consensus financial rules for the operation of this Protocol.

Article 14: Relationship of this Protocol to the Convention

Except as otherwise provided in this Protocol, the provisions of the Convention relating to its protocols shall apply to this Protocol.

Article 15: Signature

This Protocol shall be open for signature by States and by regional economic integration organizations in Montreal on 16 September 1987, in Ottawa from 17 September 1987 to 16 January 1988, and at United Nations Headquarters in New York from 17 January 1988 to 15 September 1988.

Article 16: Entry into force

1. This Protocol shall enter into force on 1 January 1989, provided that at least eleven instruments of ratification, acceptance, approval of the Protocol or accession thereto have been deposited by States or regional economic integration organizations representing at least two-thirds of 1986 estimated global consumption of the controlled substances, and the provisions of paragraph 1 of Article 17 of the Convention have been fulfilled. In the event that these conditions have not been fulfilled by that date, the Protocol shall enter into force on the ninetieth day following the date on which the conditions have been fulfilled.

2. For the purposes of paragraph 1, any such instrument deposited by a regional economic integration organization shall not be counted as additional to those deposited by member States of such organization.

3. After the entry into force of this Protocol, any State or regional economic integration organization shall become a Party to it on the ninetieth day following the date of deposit of its instrument of ratification, acceptance, approval or accession.

Article 17: Parties joining after entry into force Subject to Article 5, any State or regional economic integration organization which becomes a Party to this Protocol after the date of its entry into force, shall fulfil forthwith the sum of the obligations under Article 2, as well as under Articles 2A to 2J Articles 2A to 2I and Article 4, that apply at that date to the States and regional economic integration organizations that became Parties on the date the Protocol entered into force.

Article 18: Reservations

No reservations may be made to this Protocol.

Article 19: Withdrawal

Any Party may withdraw from this Protocol by giving written notification to the Depositary at any time after four years of assuming the obligations specified in paragraph 1 of Article 2A. Any such withdrawal shall take effect upon expiry of one year after the date of its receipt by the Depositary, or on such later date as may be specified in the notification of the withdrawal.

Article 20: Authentic texts

The original of this Protocol, of which the Arabic, Chinese, English, French, Russian and Spanish texts are equally authentic, shall be deposited with the Secretary-General of the United Nations.

IN WITNESS WHEREOF THE UNDERSIGNED, BEING DULY AUTHORIZED TO THAT EFFECT, HAVE SIGNED THIS PROTOCOL.

DONE AT MONTREAL THIS SIXTEENTH DAY OF SEPTEMBER, ONE THOUSAND NINE HUNDRED

AND EIGHTY SEVEN.

2.4 Environmental Impact Assessment (EIA)

➢ Origin of EIA

Before the First World War, rapid industrialization and urbanization in western countries was causing rapid loss of natural resources. This continued to the period after the Second World War giving rise to concerns for pollution, quality of life and environmental stress. In early 60s, investors and people realized that the projects they were under taking were affecting the environment, resources, raw materials and people. As a result of this, pressure groups formed with the aim of getting a tool that can be used to safeguard the environment in any development. The USA decided to respond to these issues and established a National Environmental Policy Act in 1970 to consider its goal in terms of environmental protection. The USA became the first country to enact legislation on EIA. This was the first time that EIA became the official tool to be used to protect the environment. The United Nations Conference on the Environment in Stockholm in 1972 and subsequent conventions formalized EIA. At present, all developed countries have environmental laws whereas most of the developing countries are still adopting it (Lee, 1995). Multilateral and bilateral lenders included EIA requirements in their project eligibility criteria (OECD, 1996).

➢ EIA in Developing Countries

Until recently, EIA as a new concept was not readily understood and accepted as a tool in developing countries. Developers resisted and argued that it was anti-development because laws and policies supporting it dictated that lands developments causing negative impacts should be discontinued. In a nutshell, EIA was considered just another bureaucratic stumbling block in the path of development. Secondly, it was conceived as a sinister means by which industrialized nations intend to keep developing countries from breaking the vicious cycle of poverty. Thirdly, the experts in the developing countries were foreigners who were viewed as agents of colonization. The need for EIAs has become increasingly important and is now a statutory requirement in many developing countries. Historically, the choice of new projects was primarily based on one criterion: economic viability. Today, a second and a third choice criteria, environmental and social impact, have become a strong yardstick, hence the triple bottom-line approach (economic, environmental and social) to project viability (Modak & Biswas, 1999).

➢ EIA: Legal, Policy & Institutional Framework

EIA takes place within the legal and/or policy and institutional frameworks established by individual countries and international agencies. EIA provision and procedure can contribute to successful implementation of project if these frameworks are adhered to.

EIA in international environmental law context Key Multilateral Environmental Agreements (MEAs) have seen review and improvements in EIA legal, policy and institutional arrangements. The key agreements are discussed below.

a) Convention on Environmental Impact Assessment in a Trans-boundary Context (Espoo, 1991). This is the first multi-lateral EIA treaty. It looks at EIA in a trans-boundary context and entered into force in 1997. The Espoo Convention sets out the obligations of Parties to assess the environmental impact of certain activities at an early stage of planning. It also lays down the general obligation of states to notify and consult each other on all major projects under consideration that are likely to have a significant adverse environmental impact across borders. Apart from stipulating responsibility of signatory countries with regards to proposals that have transboundary impacts, it describes the principles, provisions, procedures to be followed and list of activities, contents of documentation and criteria of significance that apply.

b) Rio Declaration (1992). Principle 17 of Rio Declaration on Environment and Development calls for use of EIA as a national decision making instrument to be used in assessing whether proposed activities are likely to have significant adverse impact on the environment. It also emphasized the role of competent national authority in the decision making process.

Agenda 21 sets the framework within which countries can establish their national environmental laws.

c) UN Convention on climate change and Biological Diversity (1992) cited EIA as an implementing mechanism of these conventions (article 4 and 14 respectively).

d) Doha Ministerial Declaration encourages countries to share expertise and experience with members wishing to perform environmental reviews at the national level (November, 2001).

e) UNECE (Aarhus) Convention on Access to Information, Public Participation in Decision Making and Access to Justice in Environmental Matters (1998) covers the decisions at the level of projects and plans, programs and policies and by extension, applies to EIA and SEA.

f) United Nations Conference on the Environment in Stockholm 1972.

- **Multilateral and Bilateral Financial Institutions Environmental Safeguards**

Investment banks like African Development Bank (AfDB), Asian Development Bank (ADB), European Bank for Reconstruction and Development (EBRD), European Investment Bank (EIB), Japanese Bank for International Cooperation (JBIC), World Bank (WB) have environmental safeguards to ensure that financing of projects is not only based on the precautionary principle, preventative action rather than curative treatment but sustainable development (WBCSD,2005).

Although their operational policies and requirements vary in certain respects, the development banks follow a relatively standard procedure for the preparation and approval of an EIA report. Borrowing countries are responsible for the preparation of the EIA, and this requirement possibly more than any other has influenced the introduction and development of EIA in many developing countries. The EIA should examine project alternatives and identify ways of improving project selection, siting, planning, design and implementation by preventing, minimizing, mitigating and compensating for adverse environmental impacts. Just like other banks, the World Bank has criteria for screening projects as follows:

Category A: If the project likely to have significant environmental impacts that are sensitive, diverse or unprecedented. These impacts may affect an area broader than the communities benefiting from infrastructure investments.

Category B: If the projects potential adverse environmental impacts on human populations or environmentally important areas are less adverse than those of Category A projects. These impacts are site-specific; few if any of them are irreversible; and in most cases mitigation measures can be designed more readily than for Category A projects.

Category C: If the project is likely to have minimal or no adverse environmental impacts. Once the project is assessed and determined as Category C, no further action would be required. Some examples of Category C projects include: Education (i.e. capacity-building, etc., not including school construction) Family planning (World Bank 1999) etc.

All projects financed by the Banks should also comply with the requirements of relevant multilateral environmental agreements (MEA) to which the host country is a party, including the Montreal Protocol (on ozone depleting substances), the UN Convention on Climate Change and the Kyoto Protocol (on greenhouse gas emissions) and the Aarhus Convention (on environmental information). All international organizations and bilateral agencies frequently update their procedures and it is important to obtain the current version from the organization.

➢ National legislations

National legislation may include a statutory requirement for an EIA to be done in a prescribed manner for specific development activities. Most legislation lists projects for which EIA is a mandatory requirement. The statutory requirement to carry out an EIA for specific projects will, for example, require registered experts to carry out the study, the authority with the help of lead agencies and technical committees to review the EIA and approve the project. Other national legal requirements that govern the use and protection of resources like water, fisheries, forests, wildlife, public health etc. must be identified and complied with during an EIA.

➢ Institutional Framework

EIA institutional systems vary from country-to-country and reflecting different types of governance. In some countries, either the Ministry of Environment or a designated authority or Planning Agency administers EIA. Environmental issues also involve many disciplines and many government bodies with general environmental and resource management laws. Data will therefore have to be collected and collated from a wide range of technical ministries, other government authorities and parastatals where applicable.

➢ ENVIRONMENTAL IMPACT ASSESSMENT (EIA) PROCESS

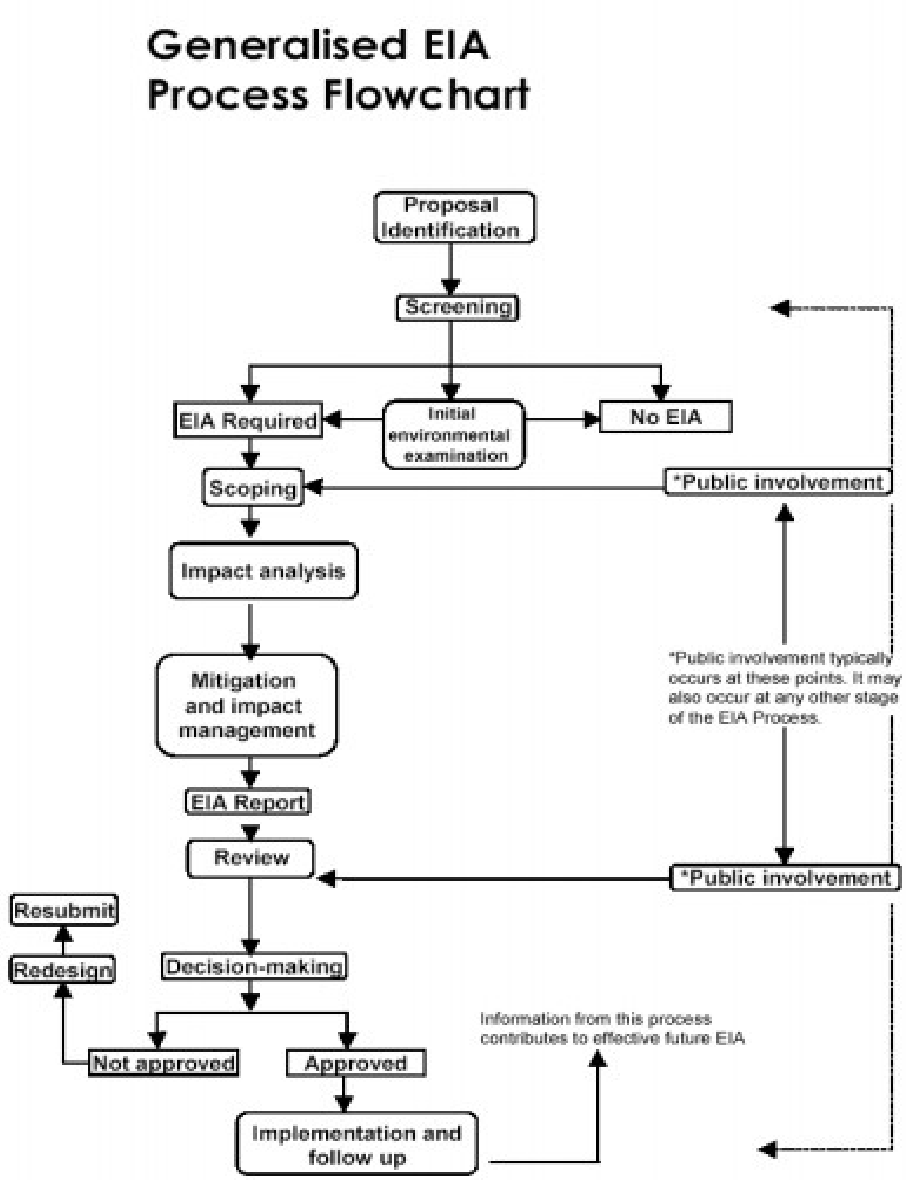
The first phase of an environmental assessment is called an Initial Environmental Examination (IEE) and the second is Environmental Impact Studies (EIS) or simply detailed EIA.

a) Initial Environmental Examination (lEE)

IEE is carried out to determine whether potentially adverse environmental effects are significant or whether mitigation measures can be adopted to reduce or eliminate these adverse effects. The IEE contains a brief statement of key environmental issues, based on readily available information, and is used in the early (pre-feasibility) phase of project planning. The IEE also suggests whether in-depth studies are needed. When an IEE is able to provide a definite solution to environmental problems, an EIA is not necessary. IEE also requires expert advice and technical input from environmental specialists so that potential environmental problems can be clearly defined.

b) Environmental Impact Assessment (EIA)

EIA is a procedure used to examine the environmental consequences or impacts, both beneficial and adverse, of a proposed development project and to ensure that these effects are taken into account in project design. The EIA is therefore based on predictions. These impacts can include all relevant aspects of the natural, social, economic and human environment. The study therefore requires a multidisciplinary approach and should be done very early at the feasibility stage of a project. In other words, a project should be assessed for its environmental feasibility. EIA should therefore be viewed as an integral part of the project planning process. Unlike the environmental audit (EA), which is conducted on existing projects, the EIA is applied to new projects and the expansion aspects of existing projects.



Screening

EIA process kicks off with project screening. Screening is done to determine whether or not a proposal should be subject to EIA and, if so, at what level of detail. Guidelines for whether or not an EIA is required are country specific depending on the laws or norms in operation. Legislation often specifies the criteria for screening and full EIA. Development banks also screen projects presented for financing to decide whether an EIA is required using their set criteria. The output of the screening process is often a document called an Initial Environmental Examination or Evaluation (IEE). The main conclusion will be a classification of the project according to its likely environmental sensitivity. This will determine whether an EIA is needed and if so, to what detail.

Scoping

The aim of EIA is not to carry out exhaustive studies on all environmental impacts for all projects. Scoping is used to identify the key issues of concern at an early stage in the planning process (Ahmed & Sammy, 1987). The results of scoping will determine the scope, depth and terms of reference to be addressed within the Environmental statement. Scoping is done to:

- Identify concerns and issues for consideration in an EIA
- Ensure a relevant EIA Enable those responsible for an EIA study to properly brief the study team on the alternatives and on impacts to be considered at different levels of analysis
- Determine the assessment methods to be used Identify all affected interests
- Provide an opportunity for public involvement in determining the factors to be assessed, and facilitate early agreement on contentious issues
- Save time and money
- Establish terms of reference (TOR) for EIA study

Scoping should be an on-going exercise throughout the course of the project.

The following environmental tools can be used in the scoping exercise -

Checklists – Checklists are standard lists of the types of impacts associated with a particular type of project. Checklists methods are primarily for organizing information or ensuring that no potential impact is overlooked. They comprise list questions on features the project and environments impacts. They are generic in nature and are used as aids in assessment.

Matrices - Matrix methods identify interactions between various project actions and environmental parameters and components. They incorporate a list of project activities with a checklist of environmental components that might be affected by these activities. A matrix of potential interactions is produced by combining these two lists (placing one on the vertical axis and the other on the horizontal axis). They should preferably cover both the construction and the operation phases of the project, because sometimes, the former causes greater impacts than the latter. However, matrices also have their disadvantages: they do not explicitly represent spatial or temporal considerations, and they do not adequately address synergistic impacts.

Networks – these are cause effect flow diagrams used to help in tracing the web relationships that exist between different activities associated with action and environmental system with which they interact.

They are also important in identifying direct and cumulative impacts. They are more complex and need expertise for their effective use.

Consultations – with decision-makers, affected communities, environmental interest groups to ensure that all potential impacts are detected. However there can be danger in this when excessive consultation is done and some unjustifiable impacts included in the ToR.

Baseline Data Collection

The term "baseline" refers to the collection of background information on the biophysical, social and economic settings proposed project area. Normally, information is obtained from secondary sources, or the acquisition of new information through field samplings, interviews, surveys and consultations with the public. The task of collecting baseline data starts right from the period of project inception; however, a majority of this task may be undertaken during scoping and actual EIA.

Baseline data is collected for two main purposes

- To provide a description of the current status and trends of environmental factors (e.g., air pollutant concentrations) of the host area against which predicted changes can be compared and evaluated in terms of significance, and

- To provide a means of detecting actual change by monitoring once a project has been initiated

Only baseline data needed to assist prediction of the impacts contained in the ToR and scoping report should be collected.

Impact Analysis and Prediction

Predicting the magnitude of a development likely impacts and evaluating their significance is core of environmental assessment process (Morris & Therivel, 1995). Prediction should be based on the available environmental baseline of the project area. Such predictions are described in quantitative or qualitative terms.

Considerations in impact prediction

Magnitude of Impact:

This is defined by the severity of each potential impact and indicates whether the impact is irreversible or, reversible and estimated potential rate of recovery. The magnitude of an impact cannot be considered high if a major adverse impact can be mitigated.

Extent of Impact:

The spatial extent or the zone of influence of the impact should always be determined. An impact can be site-specific or limited to the project area; a locally occurring impact within the locality of the proposed project; a regional impact that may extend beyond the local area and a national impact affecting resources on a national scale and sometimes trans-boundary impacts, which might be international.

Duration of Impact:

Environmental impacts have a temporal dimension and needs to be considered in an EIA. Impacts arising at different phases of the project cycle may need to be considered. An impact that generally lasts for only three to nine years after project completion may be classified as short-term. An impact, which continues for 10 to 20 years, may be defined as medium-term, and impacts that last beyond 20 years are considered as long-term.

- **Significance of the Impact**

This refers to the value or amount of the impact. Once an impact has been predicted, its significance must be evaluated using an appropriate choice of criteria. The most important forms of criterion are:

- Specific legal requirements e.g. national laws, standards, international agreements and conventions, relevant policies etc.
- Public views and complaints
- Threat to sensitive ecosystems and resources e.g. can lead to extinction of species and depletion of resources, which can result, into conflicts.
- Geographical extent of the impact e.g. has trans- boundary implications.
- Cost of mitigation Duration (time period over which they will occur)
- Likelihood or probability of occurrence (very likely, unlikely, etc.)
- Reversibility of impact (natural recovery or aided by human intervention)
- Number (and characteristics) of people likely to be affected and their locations
- Cumulative impacts e.g. adding more impacts to existing ones.
- Uncertainty in prediction due to lack of accurate data or complex systems. Several techniques can be used in predicting the impacts. The choices should be appropriate to the circumstances. These can be based on: Professional judgment with adequate reasoning and supporting data. This technique requires high professional experience. Experiments or tests. These can be expensive.
- Past experience
- Numerical calculations & mathematical models. These can require a lot of data and competency in mathematical modelling without which hidden errors can arise
- Physical or visual analysis.
- Detailed description is needed to present the impact.
- Geographical information systems,
- Risk assessment, and
- Economic valuation of environmental impacts

- **Analysis of alternatives**

Analysis of alternative is done to establish the preferred or most environmentally sound, financially feasible and benign option for achieving project objectives.

The World Bank directives requires systematic comparison of proposed investment design in terms of site, technology, processes etc. in terms of their impacts and feasibility of their mitigation, capital, recurrent costs, suitability under local conditions and institutional, training and monitoring requirements (World bank 1999). For each alternative, the environmental cost should be quantified to the extent possible and economic values attached where feasible, and the basic for selected alternative stated. The analysis of alternative should include a NO PROJECT alternative.

- **Mitigation and Impact Management**

Mitigation is done to avoid, minimize or offset predicted adverse impacts and, where appropriate, to incorporate these into an environmental management plan or system. For each potential adverse impact the plan for its mitigation at each stage of the project should be documented and costed, as this is very important in the selection of the preferred alternative. The objectives of mitigation therefore are to:

- find better alternatives and ways of doing things;
- enhance the environmental and social benefits of a project
- avoid, minimise or remedy adverse impacts; and
- ensure that residual adverse impacts are kept within acceptable levels

2.5 Environmental Audit

Environmental auditing is a management tool which simply inspects the environmental management activities performed by the industries or organizations and makes them aware of new cleaner technology. For the impact of industries and their product on natural resources and environmental quality it is necessary to have "Environmental Audit" to ensure sustainable industrial developments. Pollution now is an inevitable consequence of modern industrial technology, rapid and convenient transport and comfortable housing, but excessive pollution may interfere with man's health and his mental, social and economic wellbeing. There is now a clarion call from every nook and corner of the society that "save the nation from this menace the pollution". Governments, all over the world, have formulated laws and regulations to correct and cure the past violations of good environmental practices. The term auditing is known to us in financial accounts and records are examined.

Environmental audit is for the impact of the industries and their products on natural resources and environmental quality. It is necessary to have 'Environmental Audit' to ensure sustainable industrial developments. Environmental Audit is a pragmatic management tool, which addresses itself to help an industry or operation, to verify compliance with environmental requirements, to evaluate the effectiveness of the environmental management system, to assess risks and to identify and correct environmental hazards. It is the examination of accounts of revenues and costs of environmental and natural resources, their estimation, depreciations and natural resources, their estimation depreciations and values recorded in the books of accounts.

Environmental organization management systems and equipment are performing with the aims of:

i. Facilitating management control of environmental practices.

ii. Assessing compliance with company policies.

iii. Facilitating professional competence. The concept of environmental auditing came into being in industrialized countries such as Canada, USA, UK and Netherland during early 1970's under a different number of approaches and names depending on the company concerned such as "Environmental review". Environmental performance appraisal, Quality approach, etc. In 1972, British Petroleum (B.P.) International group first used this concept. The United States Environmental Protection Agency published their environmental policy in 1986 followed by International Chamber of Commerce book let on environmental auditing in the year1988.

An environmental audit as defined in ISO 14000 is a systematic, documented verification process of objectively obtaining and evaluating audit evidence to determine whether specified environmental activities, events, conditions, management systems, or information about these matters conform with audit criteria, and communicating the results of this process.

The International Chamber of Commerce defines environmental auditing as, "a management tool comprising a systematic documented, periodic and objective evaluation of how well environmental organization, management and equipment are performing with the aim of contributing to safeguarding the environment by facilitating management control of environmental practices and assessing compliance with company policies which could include meeting regulatory requirements. Environmental audits are generally performed on a routine or periodic basis.

More frequent assessments may be appropriate at any facility that has been targeted for more frequent federal, state/province and/or local inspections, and/or been issued a notice of violation, or subject to some form enforcement proceeding since the last assessment. The audit should be carried out following the ISO Standard 14011 on Environmental Auditing Procedures including a kick-off meeting, detailed inspection, interviews, document review as well as a closeout meeting with the plant management.

OBJECTIVES OF ENVIRONMENTAL AUDITING

1. An environmental audit programme which is designed and implemented properly can enhance an industry's environmental performance.

2. Monitoring the scale of optimum utilization of the resources and evaluating the company at national & international level.

3. To suggest for using alternative energy for the conservation of energy resources.

4. Evaluation of waste water quality and determination of waste water characteristics & their effects on the living system.

5. Classification of the categories of solid waste hazardous waste their sources, quantities & characteristics.

6. Introduction and implementation of time saving technologies in production.

7. Maintains of Labour / Occupational health & medicine.

8. Proper documentation of environmental compliance status.

9. To help in minimizing the wastes through modern cleaner technologies.

10. Regular environmental auditing once in a year will help in producing environmentally educated &technically sound personnel.

The purpose of the environmental audit is to provide an indication to the management of the improvements while environmental organization system & equipment are performing. To fulfil this purpose it is essential that audits should be seen as the responsibility of the company. The audit work can be voluntary and for the advantage of the company. The audit work can be done systematically and efficiently by the help of environmental auditing programme. It helps in the proper utilization of natural resources as a whole it improves environmental quality.

ORIGIN OF ENVIRONMENTAL AUDIT

Environmental audit is at a relatively formative stage even in western countries in the recent past with internal audit departments checking them as part of their legal compliance program. Increasing public concern about the environmental translated into legislative activism in this field. Proactive companies started demonstrating their environmental concern and the steps initiated by them in terms of sustainable development in their accountability report. The enactment and strict enforcement of various environmental laws by developed countries in respect of maintenance of air, water, waste management, transportation and storage and disposition of hazardous goods, noise etc., made the non- compliance with environmental legislation a costly affair for the business. The enforcing authorities were armed with the search and seizure power. Though environment audit is not a statutory requirement in advanced countries, business house use this as a tool of their demonstration for sustainable development and social commitment.

ENVIRONMENTAL AUDIT SCHEME AND ITS COMPONENTS

This particular tool is very important aspect of the environmental audit for the total management system in terms of its being an asset or a liability for the industry`s environmental performance. Environmental system is with a broad aim for a green environment. It helps in reducing waste.

 It helps in assessing compliance with regulatory requirement.

 It also helps in prevention control of effect of pollutant.

 It promotes relationship between qualified technician professionals, individual industries, State Pollution Control Board, other public authorities and industrial association etc.

Environmental Audit Scheme (EAS) has three following components.

1. State Pollution Control Board

2. Internal Auditor Board

3. External Auditor

1. State Pollution Control Board-It plays active role in implementing the environmental audit effectively. The steps involved in state pollution control board are mentioned. To prepare format of audit report on all the aspect of environmental protection.

- To appoint internal auditors to prepare industries audit report.
- Evaluation and verification of audit reports.
- Initiating the action on evaluated report.

2. Internal Auditor-The selection of auditor consist of experienced experts from various backgrounds. A qualified auditor should be required, as per the rules of State Pollution Control Board with well-equipped laboratory facility for analysis of water and air samples.

3. External Auditor-Team should be approved by State Pollution Control Board. Evaluated and verified reports have to send their comments to State Pollution Control Board for further action.

PRINCIPAL AREAS OF ENVIRONMENTAL AUDITING

The principal areas covered are

a. Material Audit - It mainly concentrate on the use of different raw material or natural use of resources, cost\unit, process wise consumption, wastage etc. Conservation of raw material, scientific storage & reuse of wastage material are taken in to consideration.

b. Energy Audit - It examines consumption of various forms of energy in different processes in any industry or organization. The main of audit is minimization, elimination of avoidable losses of valuable energy & their conservation.

c. Water Audit - Consumption of water at different sources is noted. It also concentrates on the reuse and recycling of water, evaluation of raw water intake, balancing of water table & other sources.

d. Health and Safety Audit - Workers &employee are the basic need of industry. Health and safety of those is well considered in audit. Proper disposal of toxic and hazardous waste, fire prevention measures etc. should be evaluated.

e. Environmental Quality Audit - Conservation of every aspect and stage of environment helps to maintain the quality. This also well noted in audit scheme.

f. Waste Audit: It covers the qualitative and quantities evaluation of waste generated from industries.

g. Engineering Audit - Use of advance technology which will cover the suitable processes and engineering application.

h. Compliance Audit: The different aspect of audits that are required to be carried out as per regulation, procedure &according to the policies of that particular industry are known as compliance audit.

ENVIRONMENTAL AUDIT ACTIVITIES

1. **Pre-audit Activities**

These include selection of the expert team and development of a plan. The plan should include defining the scope of the work, selecting priority areas, laying down the procedure & allocating team resources. There are four key activities.

- Submitting pre-visit questionnaire of the facility.
- Reviewing relevant regulation.
- Defining audit scope and team responsibilities
- Reviewing audit check lists.

2. **<u>On-site activities</u>**

Meeting of the team with the appropriate with the appropriate personnel of the unit &with discussion ofthe objective. The three primary functions on site activities are

- Record & documentation review
- Interview with staff
- Physical inspection of the facilities.

3. **<u>Post-audit Activities</u>**

Development of raw material balance analysis for each process unit of the industry highlighting analysis for each process unit of the industry highlighting the proposed utilization of raw material and ideas regarding reuse can be thought of. Water, energy consumption, vegetation removal all the sensitive parts must be balanced. Evaluation of pollution audit, emission standard are carried out.

a. Issue of draft report
b. issue of final report

ADVANTAGES

- ✓ Preparation of Environmental management plan.
- ✓ Assessment of environmental input and risks.
- ✓ Identifying areas of strength and weakness for improvements.
- ✓ Evaluation of pollution control.
- ✓ Verification of compliance with laws.
- ✓ Assuring safety of plant, environment & human beings.
- ✓ Enhancement of loss prevention, manpower development and marketing.
- ✓ Budgeting for pollution control, waste prevention, reduction, recycling and reuse.
- ✓ Providing an opportunity for management to give credit for good environmental performance.
- ✓ As a whole environmental audit plays an important role in minimizing the environmental problem locally, regionally, nationally and internationally.

Why Implement an Environmental Auditing Program?

An environmental audit determines how well your business complies with environmental laws and regulations. Environmental audits are undertaken for a variety of reasons. An audit may relate to a strict compliance audit, where facility activities are reviewed against legislative requirements, or as part of a company's management system to ensure environmental best practice. Audits may cover a wide spectrum of environmental issues or be focused on a particular aspect such as air and water permits. There are potentially significant liabilities associated with auditing, therefore auditors must be experienced and fully conversant with environmental auditing requirements.

Audits have always been an important business tool, but now there are even more incentives to perform an environmental audit, especially in the aggregate industry. With increased awareness of the need for environment protection, the aggregate industry will need to rely increasingly on environmental audits. The principal aims are to ensure compliance with regulatory agencies, as well as to identify and evaluate potential liabilities, risks and hazards. This, in turn, will assist in assessing the viability of an operation with the inclusion of costs associated with reducing environmental risks and liabilities to acceptable levels. In general, the benefits of a successful environmental auditing program can be categorized as improving a company's financial position, compliance status, and stakeholder relations.

Below are examples of benefits that a successful environmental auditing program could generate:

Financial

- ✓ Helps avoid fines by regulatory agencies
- ✓ Identifies issues of non-compliance sooner versus later allowing for proactive financial planning
- ✓ Lowers corrective action costs
- ✓ Waste minimization opportunities realized, leading to reduced operating costs

Compliance

- ✓ Reduced agency enforcement actions and penalties
- ✓ Increased employee awareness of environmental standards and responsibilities

Stakeholder Relations

- ✓ Improved employee relations and increased morale
- ✓ Improved community image of the Company
- ✓ Goodwill
- ✓ Firms with sound environmental stewardship programs are desirable to investors

2.6 Environmental Management Planning

The Environment Management Plan (EMP) would consist of all mitigation measures for each component of the environment due to the activities increased during the construction, operation and the entire life cycle to minimize adverse environmental impacts resulting from the activities of the project. It would also delineate the environmental monitoring plan for compliance of various environmental regulations. It will state the steps to be taken in case of emergency such as accidents at the sites including fire. The detailed EMP for the complex is given below.

Environmental Management Plan

The Environment Management Plan (EMP) is a site specific plan developed to ensure that the project is implemented in an environmental sustainable manner where all contractors and subcontractors, including consultants, understand the potential environmental risks arising from the project and take appropriate actions to properly manage that risk. EMP also ensures that the project implementation is carried out in accordance with the design by taking appropriate mitigation actions to reduce adverse environmental impacts during its life cycle. The plan outlines existing and potential problems that may adversely impact the environment and recommends corrective measures where required. Also, the plan outlines roles and responsibility of the key personnel and contractors who will be in-charge of the responsibilities to manage the project site.

The EMP is generally

- Prepared in accordance with rules and requirements of the MoEF and CPCB/ SPCB

- To ensure that the component of facility are operated in accordance with the design

- A process that confirms proper operation through supervision and monitoring

- A system that addresses public complaints during construction and operation of the facilities and

- A plan that ensures remedial measures is implemented immediately. The key benefits of the EMP are that it offers means of managing its environmental performance thereby allowing it to contribute to improved environmental quality. The other benefits include cost control and improved relations with the stakeholders.

Objectives of the EMP

The objectives of the EMP are to:

- Identify a range of mitigation measures which could reduce and mitigate the potential impacts to minimal or insignificant levels;

- To identify measures that could optimize beneficial impacts;

- To create management structures that address the concerns and complaints of stakeholders with regards to the development;

- To establish a method of monitoring and auditing environmental management practices during all phases of development;

- Ensure that the construction and operational phases of the project continues within the principles of Integrated Environmental Management;

- Detail specific actions deemed necessary to assist in mitigating the environmental impact of the project;

- Ensure that the safety recommendations are complied with;

- Propose mechanisms for monitoring compliance with the EMP and reporting thereon; and

- Specify time periods within which the measures contemplated in the final environmental management plan must be implemented, where appropriate.

EMP Roles and Responsibilities

Several professionals will form part of the construction team. The most important from an environmental perspective are the Project Manager (Galetech Energy Developments of Ireland), the Project HSE Officer, the EPC Contractor, and the developer (Kipeto Energy Limited).

The Project Manager is responsible for ensuring that the EMP is implemented during the pre-construction and construction phases of the project. The Project HSE Officer is responsible for monitoring the implementation of the EMP during the design, pre-construction and construction phases of the project.

The EPC contractor is responsible for abiding by the mitigation measures of the EMP which are implemented by the Project Manager during the construction phase.

The Project Manager is responsible for ensuring that the EPC contractor complies with the mitigation measures and EMP requirements during the design, preconstruction and construction phases of the project. An Operations and Maintenance (O&M) company will be responsible for implementation of the EMP during the operational and decommissioning phases of the project.

Decommissioning will however entail the appointment of a new professional team and responsibilities will be similar to those during the design, pre-construction and construction phases. It is unlikely that the transmission line will be decommissioned for several years.

Environmental Monitoring

A monitoring program will be implemented for the duration of the construction phase of the project. This program will include:

- Monthly environmental inspections to confirm compliance with the EMP and EIA License conditions. These inspections can be conducted randomly and do not require prior arrangement with the project manager;

- Compilation of an inspection report complete with corrective actions for implementation;

- Monthly HSE committee meetings to be held to ensure compliance with the OSHA and its subsidiary legislation.

The HSE Officer shall keep a photographic record of any damage to areas outside the demarcated site area. The date, time of damage, type of damage and reason for the damage shall be recorded in full to ensure the responsible party is held liable. All claims for compensation emanating from damage should be directed to the Project HSE Officer for appraisal. The Contractor shall be held liable for all unnecessary damage to the environment. A register shall be kept of all complaints from the Landowner or community. All complaints/claims shall be handled in a timely manner to ensure timeous rectification/payment by the responsible party.

It should be noted that it is difficult to outline a formal monitoring protocol for specific environmental parameters and key impacts until detailed transmission line design have been completed. A formal monitoring protocol will be included within the revised EMP once the detailed transmission line design has been completed, and once recommendations and conditions from the reviewing authority have been received. It should further be noted that for the same reasons mentioned above, it is difficult to delineate the cost of the EMP for specific mitigation measures and therefore this has been excluded from the EMP tables.

The EPC contractor shall be responsible for acquiring all necessary permits during the construction phase of the project. Such licenses include any abstraction of water permits, local authority approvals for camp site locations and operations, extraction of aggregates from borrow pits and their rehabilitation, etc.

EMP Requirements for Pre-Construction Phase

The requirements that need to be fulfilled during the pre-construction phase of the project are as follows:

- There should be continuous liaison between the Proponent, the EPC contractor and Landowners along the wayleave to ensure all parties are appropriately informed of construction phase activities at all times;

- The Landowners should be informed of the starting date of construction as well as the phases in which the construction will take place;

- The EPC contractor must adhere to all conditions of contract including the Environmental Management Plan;

- The EPC contractor should plan the construction program taking cognizance of climatic conditions along the wayleave especially wet seasons and disruptions that can be caused by heavy rains;

- Where existing private roads are in a bad state of repair, such roads' condition shall be documented before they are used for construction purposes. This will allow for easy assessment of any damage to the roads which may result from the construction process. If necessary some repairs should be done to prevent damage to equipment;

- The construction site office must keep a proper record of all complaints received and actions taken to resolve the complaints;

- A Project HSE Officer should be appointed by the Proponent and Contractor HSE officer should be appointed by the EPC contractor to implement this EMP as well as deal with Landowner related matters;

- Internal and external environmental inspections and audits should be undertaken during and upon completion of construction. The frequency of these audits should be quarterly;

- The Project HSE Officer should conduct regular inspections along the wayleave in order to maintain good control over the construction process during the construction phase;

- A formal communications protocol should be set up during this phase. The aim of the protocol should be to ensure that effective communication on key issues that may arise during construction be maintained between key parties such as the Project HSE Officer, project manager and EPC contractor. The protocol should ensure that concerns/issues raised by stakeholders are formally recorded and considered and where necessary acted upon. If necessary, a forum for communicating with key stakeholders on a regular basis may need to be set up. This could be done through the EPC contractor's site office that would meet on a regular basis. The communications protocol should be maintained throughout the construction phase.

ENVIRONMENT MANAGEMENT PLAN

An Environmental Management Plan (EMP) will be required to mitigate the predicted adverse environmental impacts during construction and operation phase of the project and these are discussed in later subsections. EMP for Air Environment Construction Phase To mitigate the impacts of PM10 & PM2.5 during the construction phase of the project, the following measures are recommended for implementation:

- A dust control plan
- Procedural changes to construction activities

Dust Control Plan

The most cost-effective dust suppressant is water because water is easily available on construction site. Water can be applied using water trucks, handled sprayers and automatic sprinkler systems. Furthermore, incoming loads could be covered to avoid loss of material in transport, especially if material is transported off-site.

EMP FOR NOISE ENVIRONMENT

Construction Phase

To mitigate the impacts of noise from construction equipment during the construction phase on the site, the following measures are recommended for implementation. Time of Operation: Noisy construction equipment has not been be allowed to use at night time. Job Rotation and Hearing Protection: Workers employed in high noise areas are not employed on shift basis. Hearing protection such as earplugs/muffs will be provided to those working very close to the noise generating machinery. Operation Phase To mitigate the impacts of noise from diesel generator set during operational phase, the following measures are recommended:

- Adoption of Noise emission control technologies
- Greenbelt development

Noise Emission Control Technologies

Source of noise in the operational phase will be from backup DG sets (which will be in operation only during power failure) and pumps & motors. All the machinery will be of highest standard of reputed make and will comply with standard i.e. The DG set room will be provided with acoustic enclosure to have minimum 75 dB(A) insertion loss or for meeting the ambient noise standard whichever is on higher side as per E (P) Act, GSR 371 (E) and its amendments.

Greenbelt Development

The following species can be used, as in a greenbelt, to serve as noise breakers:

- ✓ Acacia auriculiformis
- ✓ Anona squamosa
- ✓ Acacia farnesiana
- ✓ Acacia mearnsii
- ✓ Acacia nilotica
- ✓ Achras sapota

EMP FOR WATER ENVIRONMENT

Construction Phase

To prevent degradation and to maintain the quality of the water source, adequate control measures have been proposed. To check the surface run-off as well as uncontrolled flow of water into any water body check dams with silt basins are proposed. The following management measures are suggested to protect the water source being polluted during the construction phase:

- Avoid excavation during monsoon season

- Care has been taken to avoid soil erosion

- Common toilets has been constructed on site during construction phase and the sullage would be channelized to the septic tanks in order to prevent sullage to enter into the water bodies

- To prevent surface and ground water contamination by oil and grease, leak-proof containers has been used for storage and transportation of oil and grease. The floors of oil and grease handling area has been kept effectively impervious. Any wash off from the oil and grease handling area or workshop has been drained through imperious drains

- Collection and settling of storm water, prohibition of equipment wash downs and prevention of soil loss and toxic release from the construction site are necessary measure to be taken to minimize water pollution

- All stacking and loading area has been provided with proper garland drains, equipped with baffles, to prevent run off from the site, to enter into any water body.

Operation Phase

In the operation phase of the project, water conservation and development measures will be taken, including all possible potential for rain water harvesting. Following measures will be adopted:

- Water source development.
- Minimizing water consumption.
- Promoting reuse of water after treatment and development of closed loop systems for different water streams.

Water Source Development

Water source development shall be practiced by installation of scientifically designed Rain Water Harvesting system. Rainwater harvesting promotes self-sufficiency and fosters an appreciation for water as a resource.

Domestic and Commercial Usage

- Use of water efficient plumbing fixtures (ultra-low flow toilets and urinals). Water efficient plumbing fixtures uses less water with no marked reduction in quality and service

- Leak detection and repair techniques.

- Sweep with a broom and pan where possible, rather than hose down for external areas.

- Meter water usage: Implies measurement and verification methods. Monitoring of water uses is a precursor for management.

Horticulture

- Drip irrigation system shall be used for the lawns and other green area. Drip irrigation can save 15-40% of the water, compared with other watering techniques.

- Plants with similar water requirements shall be grouped on common zones to match precipitation heads and emitters.

- Use of low-angle sprinklers for lawn areas.

- Select controllers with adjustable watering schedules and moisture sensors to account for seasonal variations and calibrate them during commissioning.

- Place 3 to 5 inches of mulch on planting beds to minimize evaporation.

Promoting Reuse of Water after Treatment and Development of Closed Loop Systems

To promote reuse of sullage and development of closed loop system for sullage segregation. Two water conservation schemes are suggested, namely:

1) Storm Water Harvest: Storm water harvest as discussed in earlier, will be utilized for artificial recharge of ground water sources

2) Wastewater recycling: Wastewater will be reused on site after treatment. Treated wastewater will be used for landscaping, flushing purpose. Following section discuss the scheme of wastewater treatment.

Storm Water Management

Contamination of Storm Water is possible from the following sources:

- Diesel and oil spills in the diesel power generator and fuel storage area
- Waste spills in the solid / hazardous waste storage area
- Oil spills and leaks in vehicle parking lots
- Silts from soil erosion in gardens
- Spillage of sludge from sludge drying area of sewage treatment plant

A detailed storm water management plan will be developed which will consider the possible impacts from above sources. The plan will incorporate best management practices which will include following:

- Regular inspection and cleaning of storm drains
- Clarifiers or oil/separators will be installed in all the parking areas. Oil / grease separators installed around parking areas and garages will be sized according to peak flow guidelines. Both clarifiers and oil/water separators will be periodically pumped in order to keep discharges within limits
- Covered waste storage areas
- Avoid application of pesticides and herbicides before wet season
- Secondary containment and dykes in fuel/oil storage facilities
- Conducting routine inspection to ensure cleanliness

- Provision of slit traps in storm water drains
- Good housekeeping in the above areas

EMP FOR LAND ENVIRONMENT

Construction Phase

The waste generated from construction activity includes construction debris, biomass from land clearing activities, waste from the temporary make shift tents for the labours and hazardous waste. Following section discuss the management of each type of waste. Besides waste generation, management of the topsoil is an important area for which management measures are required.

Construction Debris

Construction debris is bulky and heavy and re-utilization and recycling is an important strategy for management of such waste. As concrete and masonry constitute the majority of waste generated, recycling of this waste by conversion to aggregate can offer benefits of reduced landfill space and reduced extraction of raw material for new construction activity. This is particularly applicable to the project site as the construction is to be completed in a phased manner. Mixed debris with high gypsum, plaster, has not been be used as fill, as they are highly susceptible to contamination, and will be send to designated solid waste landfill site. Metal scrap from structural steel, piping, concrete reinforcement and sheet metal work has been removed from the site by construction contractors. A significant portion of wood scrap has been reused on site. Recyclable wastes such as plastics, glass fiber insulation, roofing etc. shall be sold to recyclers.

Hazardous waste

Construction sites are sources of many toxic substances such as paints, solvents wood preservatives, pesticides, adhesives and sealants. Hazardous waste generated during construction phase shall be stored in sealed containers and disposed off as per The Hazardous Wastes (Management, Handling & Transboundary Movement) Rules, 2008. Some management practices to be developed are:

- Herbicides and pesticide has not been over applied (small-scale applications) and not applied prior to rain.
- Paintbrushes and equipment for water and oil based paints shall be cleaned within a contained area and has not been allowed to contaminate site soils, water courses or drainage systems.
- Provision of adequate hazardous waste storage facilities. Hazardous waste collection containers has been located as per safety norms and designated hazardous waste storage areas will be away from storm drains or watercourses
- Segregation of potentially hazardous waste from non-hazardous construction site debris
- Well labelled all hazardous waste containers with the waste being stored and the date of generation
- Instruct employees and subcontractors in identification of hazardous and solid waste

Even with careful management, some of these substances are released into air, soil and water and many are hazardous to workers. With these reasons, the best choice is to avoid their use as much as possible by using low-toxicity substitutes and low VOC (Volatile Organic Compound) materials.

Waste from Temporary Makes Shift Tents for Labors

Wastes generated from temporary make shift labor tents have mainly comprise of household domestic waste, which will be managed by the contractor of the site. The sullage generated will be channelized to the septic tank.

Top Soil Management

To minimize disruption of soil and for conservation of top soil, the contractor has keept the top soil cover separately and stockpile it. After the construction activity is over, top soil will be utilized for landscaping activity. Other measures, which would be followed to prevent soil erosion and contamination include:

- Maximize use of organic fertilizer for landscaping and green belt development.

- To prevent soil contamination by oil/grease, leaf proof containers has been used for storage and transportation of oil/grease and wash off from the oil/grease handling area has been drained through impervious drains and treated appropriately before disposal.

- Removal of as little vegetation as possible during the development and re-vegetation of bare areas after the project.

- Working in a small area at a point of time (phase wise construction)

- Construction of erosion prevention troughs/berms.

Operational Phase

The philosophy of solid waste management at the complex will be to encouraging the four R's of waste i.e. Reduction, Reuse, Recycling and Recovery (materials & energy). Regular public awareness meetings will be conducted to involve the people in the proper segregation and storage techniques. The Environmental Management Plan for the solid waste focuses on three major components during the life cycle of the waste management system i.e., collection and transportation, treatment or disposal and closure and post-closure care of treatment/disposal facility.

Collection and Transportation

- During the collection stage, the solid waste of the project will be segregated into biodegradable waste and non-biodegradable. Biodegradable waste and non biodegradable waste will be collected in separate bins. Biodegradable waste will be treated in organic waste converterThe recyclable wastes will be sent off to recyclabers. Proper guidelines for segregation, collection and storage will be prepared as per MSW Rules, 2000.

- To minimize littering and odour, waste will be stored in well-designed containers/ bins that will be located at strategic locations to minimize disturbance in traffic flow

- Care would be taken such that the collection vehicles are well maintained and generate minimum noise and emissions. During transportation of the waste, it will be covered to avoid littering.

Disposal

With regards to the disposal/treatment of waste, the management will take the services of the authorized agency for waste management and disposal of the same on the project site during its operational phase.

Unit-14

2.7 Concept and Methods of Alternative Agriculture

Sustainability and food security are the major challenges faced by third world countries for the past several decades. Most of the third world countries are also facing problems of climate change, increasing population, overexploitation of natural resources and resource degradation associated with rapid economic growth. Among the scientific and policy circles there are controversies in using inorganic chemicals and biotechnology for sustaining the agricultural production.

Evolution of Alternative Farming Strategies

The interest in the sustainability of agricultural and food systems can be traced to environmental concerns that began to appear in the 1950s–1960s. However, ideas about sustainability date back at least to the oldest surviving writings from India, China, Greece and Rome (Pretty 2006;Dhama et al. 2005;Sofia et al. 2006;Balasubramanian et al. 2009).Today there is an urgent need of sustainable agricultural technologies and practices that (1) do not have adverse effects on the environment i.e. partly because the environment is an important asset for farming (2) are accessible to and effective for farmers (3) lead to both improvements in food productivity and have positive side effects on environmental goods and services. Sustainability in agricultural systems incorporates concepts of both resilience i.e. the capacity of systems to buffer shocks and stresses, and persistence i.e. the capacity of systems to continue over long periods. This culminates in many wider economic, social and environmental outcomes (Pretty 2008; Bradford and Wichner 2009).

Scope of Alternative Farming Practices

The majority of sustainable alternatives farming techniques indicate a return of time tested traditional and eco-friendly practices. The ultimate goals of sustainable agriculture stated by several researches like Vandermeer (1998), Xu et al. (2000), Greene and Kremen (2003) are: to

- maintain or improve the natural resource base
- protect the environment
- ensure profitability to the farmers
- conserve energy
- increase the farm productivity
- improve food quality, security and safety and
- create more viable and vibrant socio-economic infrastructure for farms and rural communities.

As Green revolution agriculture is detrimental to the environment as well as the local communities, there is an urgent need to find more sustainable farming methods (Mader et al. 2002; Thamaga-Chitja and Sheryl 2008).

Urgent Need for Adopting Sustainable Alternative Farming Techniques

The Green Revolution agriculture aimed at stimulating agriculture production primarily by replacing traditional varieties of crops by high response varieties-the hybrids. Green revolution agriculture has vast and adverse impacts on the earth. During the late 1970s efforts were initiated to bring the traditional practices into the realm of modern agricultural science (Bene et al. 1977;LaSalle et al. 2008). These initiatives arose from the frustrations arising from failure of the green revolution (Evenson and Gollin 2003;FAO 2007) to benefit poor farmers and those in less-productive agro-ecological environments. More than two-thirds of human water useis for agriculture and crop and livestock production. They are the main source of water pollution by nitrates, phosphates and pesticides thus impacting agriculture, forestry and fisheries leading to loss of the world's biodiversity (Robertson and Swinton 2005;Nair 2008). The sustainable alternative organic farming practices play a vital role in maintaining biodiversity, regulating climate change and energy use, conserving ecosystem services, optimizing productivity and ensuring food security.

Types of Alternative Farming Techniques

Environmental sustainability is a major driving force for the development and adoption of sustainable farming practices where monoculture production of agriculture and forestry commodities has led to reduced biodiversity and loss of wildlife habitat, increased non-point source pollution of ground and surface water, and deterioration of family farms (Jose and Gordon 2008;Rigueiro et al. 2008).The different forms of integrated land-use systems that embrace the concepts of sustainable agriculture include (Fukuoka 1985;Sachchidananda and Rajiv 1999):

- Organic farming
- Bio-dynamic farming
- No tillage farming
- Urban and Peri-urban Farming
- Natural farming
- Eco-farming
- Permaculture
- Polyculture
- Integrated farming system
- Floating Farming

These are the predominant potential sustainable farming techniques practiced in various parts of the world. This article reviews the potentials, constraints, strategies and case studies for these ten alternative farming techniques. Based on an extensive critical review of literature, it may be concluded that these farming techniques have demonstrated their ecological, economic, social and cultural sustainability. The last section for the review comprises of some innovative endogenous farming techniques practices in India, as there is a rapid revival of alternative farming in recent decades.

Organic Farming

The principles of organic farming is the maintenance of soil fertility by bio-intensive nutrient management, recycling of agricultural wastes, vermicomposting, avoidance or reduction of external inputs, use of natural forms of pest management and weed control (Goldsmith and Hildeyard 1996;Hansen et al. 2006). The organic movement began in the 1930s and 1940s as a reaction to the growing reliance of agriculture on synthetic fertilizers. Organic farming is a form of agriculture which excludes the use of synthetic fertilizers and pesticides; plant growth regulators, livestock feed additives, and genetically modified organisms. Organic agriculture can be considered a subset of sustainable agriculture, the difference being that organic implies certification in accordance with legal standards. Sir Albert Howard was widely considered to be the father of modern organic farming worked as an agricultural adviser in Pusa, Bengal from 1905 to 1924.He documented traditional Indian farming practices and came to regard them as superior to conventional agriculture, Rudolf Steiner a German philosopher made influential strides in the earliest organic theory with his biodynamic agriculture.

Organic agriculture world over involves certain basic steps as like:

- Green manuring
- Bio fertilizers
- Crop rotation
- Cover cropping
- Soil Health Management

Green Manuring

A green manure is a type of cover crop grown primarily to add nutrients and organic matter to the soil for soil improvement and soil protection. Typically a green manure crop is grown for a specific period, plowed and incorporated into the soil. Leguminous green manures contain nitrogen-fixing symbiotic bacteria in root nodules that fix atmospheric nitrogen in a form that plants can use.

• Green manures increase the percentage of organic matter (biomass) in the soil, thereby improving water retention, aeration, and other soil characteristics.

• The root systems of some varieties of green manure grow deep in the soil and bring up nutrient resources unavailable to shallower-rooted crops.

• Common cover crop functions of weed suppression and prevention of soil erosion and compaction are often also taken into account when selecting and using green manures.

• Some green manure crops, when allowed to flower, provide forage for pollinating insects.

The green manure crops could contribute 30–60 kg nitrogen per hectare annuallyto the subsequent crop and is an inexpensive source of organic fertilizer to buildup or maintain soil fertility (Amanullah 2008). For instance, the rice yield could be significantly improved by incorporating green manure and stem nodulating green manure has the capacity to fix approximately 150–220 kg N ha 1in 50–60days (Shivay and Rahal 2008). Green manuring alone (without fertilizer nitrogen)manifested an yield increase of toria by122% equivalent to solitary application of60kg N ha 1and the residual effect of green manuring on the following sunflower crop resulted in an additional yield of317kg ha 1(Bahi and Pasricha 2001). Organic agriculture is no longer a phenomenon of developed countries. It is now commercially practiced in 120 countries, representing 31 million ha of certified croplands and pastures (0.7% of global agricultural lands and an average of 4% in the European Union), 62 million ha of certified wild lands (for organic collection of bamboo shoots, wild berries, mushrooms and nuts) and a market of US$40 billion in 2006 (2% of food retail in developed countries) (Willer and Youssefi 2007).Although difficult to quantify, non-certified organic systems e.g. indigenous models that follow organic principles by intent or by default) of several million small farmers may represent at least an equivalent share in subsistence agriculture of developing countries (FAO 2007).

Bio-fertilizers

Bio-fertilizers are the substance which contains symbiotic nutrients fixating living microbes which are capable of colonizing in rhizosphere and enhances plant growth by increasing the availability of primary nutrients or by synthesizing growth promoting. The plant inoculation with Azospirillum promoted the uptake of KC, NO 3and H2PO 4, releases various metabolites such as auxines, cytokines, riboflavin and vitamins leading to higher growth in various legume and non-leguminous plant(Saubidet et al. 2000;Matriu and Dakora 2004). Azospirillum and Pseudomonas fluorescens colonize plant roots and exert beneficial effects on plant growth and development (Bashan et al. 2004;Choonget al. 2005). Rhizobium, Azospirillum and phosphobacteria encourage plant growth by producing growth regulators, facilitating nutrient uptake, accelerating miner-alization, reducing plant stress, stimulating nodulation and promoting nitrogenfixation (Somers et al. 2005;Remans et al. 2006;Mallik and Williams 2008).

Crop Rotation

Crop rotations serve to provide new above-and below-ground habitats as each new crop has a distinct chemical and biological make-up, introducing new vegetation types to the landscape eventually increases crop residues to the soil ecosystem (Magdoff and Harold 2000). Different crop residues promote or inhibit different soil organisms which may have inhibitory or growth promoting effects to subsequent crops. By interrupting the continuous presence of a crop host, crop rotation serves to break the build-up in the cycles of weeds and insects and diseases, thus eliminating the need for pesticide application. Fallow periods i.e. ground left uncultivated for an extended period of time, allow a limited amount of secondary succession to advance and hence the recovery of the diversity of both terrestrial and below-ground species are possible (Ditsch et al. 1993; Svotwa et al. 2008; LaSalle et al. 2008).

Cover Cropping

Cover cropping is an ideal cropping pattern adopted specifically for soil improvement purposes. Both annual and perennial cover crops used to harnessing natural resources effectively in above-and below-ground biodiversity (Burgos and Talbert 1996;Anonymous 2001). Cover crops may provide a physical temporary habitat for many different species of ground-nesting birds, small mammals as well as nectar and pollen sources for many species of insects. The habitat value of cover crops varies by species and variety therefore cover crops must be carefully selected to meet specific management objectives. Cover crops root system improves water penetration and prevents soil erosion (Roberson et al. 1991;Tyler et al. 1994;Sainjuand Singh 1997).

Cereal cover prevents, excessive water consumption, nutrient leaching into sensitive water ways and can be an important source of organic matter when incorporated into the soil (Holderbaum et al. 1990; Sullivan et al. 1991). The use of perennial cover crops in farms is an effective means of enhancing the biodiversity and productive capacity of cropping systems by minimizing the environmental risks associated with chemical use (Creamer and Bennett 1997; Costello 1999).

Soil Health Management

Enhancing soil quality is essential for maintaining agricultural productivity and minimizing environmental degradation. Organic farming plays a key role in maintaining soil quality (Lal 2008). Intensive chemical agricultural practices either depletes soil nutrients or resulting in over-reliance of inorganic fertilizers leading to nutrient build up can be harmful to yields and the environment (Craswell 1998;Limpinuntana et al. 2001;Noble and Ruaysoongnern 2002). The use of naturally occurring soil mineral amendments e.g. rock phosphate, sulphate of potash, serves to supply essential plant nutrients and reduces nutrient leaching and runoff. Compost is used to improve and maintain soil organic matter levels. The Higher soil biodiversity in organic farms has shown to increase the rate of nutrient cycling, improve soil aggregation and aggregate stability and improve thedisease suppression of agricultural soils (Tugel et al. 2000; Mader 2002; Fulleret al. 2005; Bhaskaran et al. 2009). Additionally, in order to work effectively, manures, vermin compost and enriched phospho-composts must be incorporated into the soil. This provides the double benefit of increasing availability for crop use and decreasing potential for runoff (Mader 2002; Bhaskaran et al. 2009).

Integrated Pest Management

Since 1960s, chemical pesticides have been the dominant approach in controlling and eliminating pests resulting in more consistent crop yields as well as a reduction in labour needed to manage the crops. Crops grown in healthy soils tend to be more resistant and resilient to pest and pathogens which require little to no applications of pest control materials (Flint 1998). Integrated Pest Management addresses both the concern of the farmer by regarding the increasing amounts of pesticide to maintain the same effectiveness on insects and human as well as ecosystem health (Anonymous 2001). However, there is an imminent danger that it can unwittingly promote the corporate interest by unnecessarily emphasizing the use of avoidable pesticides and fungicides against the locally available more economically, ecologically and culturally appropriate pest control solutions. For instance the effectiveness and relevance of *panchagavya*, biofertilisers and bio-pesticides have been rediscovered recently by the scientific community (Dhamaet al. 2005;Sofia et al. 2006;Babou et al. 2009;Balasubramanian et al. 2009).

Bio-Dynamic Farming

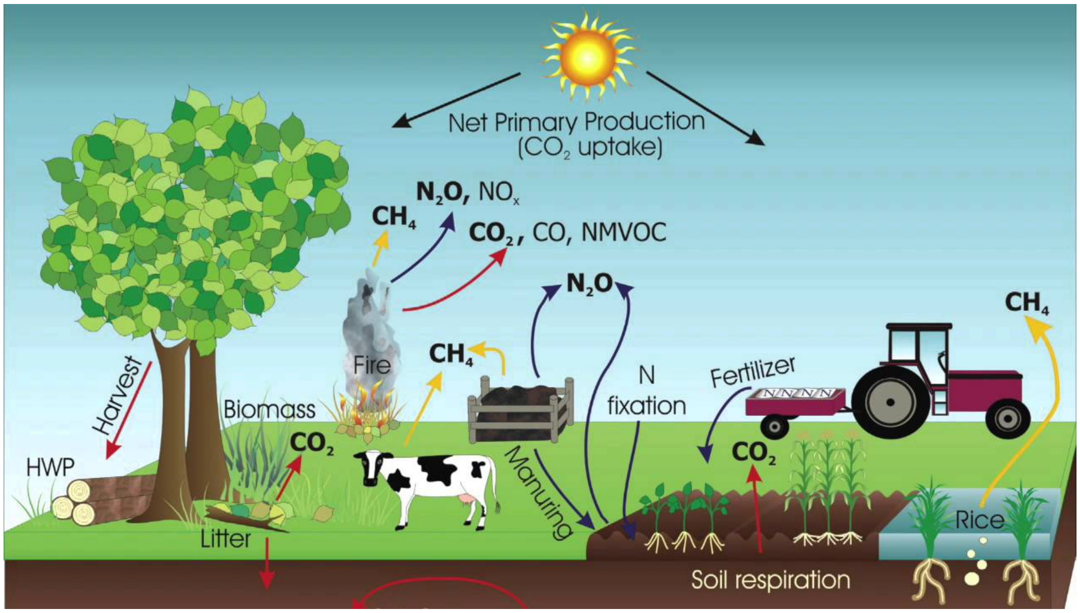
Biodynamic agriculture was the first ecological farming systems arise in response to commercial fertilizers and specialized agriculture after the turn of the century yet it remains largely unknown to the modern farmer. Biodynamic farming places great importance on the rhythmic positions of moon, sun and planets when sowing seeds, transplanting, applying liquid manures or spraying fruit trees and crops.

Biodynamic (BD) agriculture is an advanced organic farming system which gains increased attention of farmers and consumers for its emphasis on food quality and soil health. There are about 4,200 Biodynamic certified farms in 43 countries over128,000 ha, according to Demeter standards (Demeter 2008;Turinek et al. 2009).Biodynamic agriculture developed out of eight lectures on agriculture given in 1924by Rudolf Steiner(1861–1925) an Austrian scientist and philosopher to a group of farmers near Breslau (which was then in the eastern part of Germany and is now Wroclaw in Poland). Biodynamic farming is a combination of biological and dynamic practices; it also involves animal manures, crop rotations, care for animal welfare, looking at the farm entity and local distribution systems. BD farming practices are also gaining importance in the face of increasing climate change, energy scarcity and population growth, where they indicate a more resilient, diverse and efficient system (Turinek et al. 2009).

No Tillage Farming

Tillage is a critical soil management practice used for weed control, seed bed preparation, root growth stimulation, soil moisture control, soil temperature control, soil compaction alleviation and incorporation of crop residues and manure (Lal 1997;Basamba et al. 2006). Masanobu Fukuoka wrote a book “The One-Straw Revolution” was one of the pioneers work about No tillage farming or Fukuoka Farming. Producing crops usually undergo regular tilling that agitates the soil in various ways and it’s usually done by tractor-drawn implements. Usually this tilling practice affects the soil compaction, loss of organic matter, degradation of soil aggregates, death or disruption of soil microbes including mycorrhiza, arthropods, and earthworms (Chan 2001;Hussain et al. 1999) and soil erosion.

No-till farming avoids these effects by excluding the use of tillage practice. By this way of farming, crop residues or other organic amenities are retained on the soil surface, sowing and fertilizing is done with minimal soil disturbance (Chan 2001).It sometimes involves in problems like residue management, increased weed and disease infestations and this can be avoided by crop rotations and cover crops(Fukuoka 1978). No-tillage farming is effective in terms of economic and soil erosion control, than any other cost effective practices which are commonly used(King 1983;Sorrenson and Montoya 1984).Tebr ̈ugge and B ̈ohrnsen(1997) reported that No-tillage is a very profitable cultivation system compared to conventional tillage because of the cost savings from lower machinery costs and lower operation costs



Unit-15

2.8 Use and Misuse of Forest Resources and Forest Conservation

Introduction

A forest is a complex ecosystem which is predominantly composed of trees, shrubs and is usually a closed canopy. Forests are storehouses of a large variety of life forms such as plants, mammals, birds, insects and reptiles etc. Also the forests have abundant microorganisms and fungi, which do the important work of decomposing dead organic matter thereby enriching the soil. Nearly 4 billion hectares of forest cover the earth's surface, roughly 30 per cent of its total land area The forest ecosystem has two components- the non-living (abiotic) and the living (biotic) component. Climate, soil type are part of the non-living component and the living component includes plants, animals and other life forms. Plants include the trees, shrubs, climbers, grasses and herbs in the forest.

Depending on the physical, geographical, climatic and ecological factors, there are different types of forest like evergreen forest (mainly composed of evergreen tree species i.e. species having leaves all throughout the year) and deciduous forest (mainly composed of deciduous tree species i.e. species having leaf-fall during particular months of the year). Each forest type forms a habitat for a specific community of animals that are adapted to live in it. The term forest implies 'natural vegetation' of the area, existing from thousands of years and supporting a variety of biodiversity, forming a complex ecosystem. Plantation is different from natural forest as these planted species are often of same type and doesn't support a variety of natural biodiversity. Forests provide various natural services and products. Many forest products are used in day-today life. Besides these, forests play important role in maintaining ecological balance & contributes to economy also.

Ecological Role of Forest

• Forests provide an environment for many species of plants and animals thus protects and sustains the diversity of nature.

• Plants provide habitat to different types of organisms. Birds build their nests on the branches of trees, animals and birds live in the hollows, insects and other organisms live in various parts of the plant.

• Forests act as hydrologic flow modulators

• Plants provide a protective canopy that lessens the impact of raindrops on the soil, thereby reducing soil erosion. Roots help to hold the soil in place. They provide shade which prevents the soil to become too dry. Thus increases the soil moisture holding capacity.

• Forests help in maintaining microclimate of the area.

• Plants clean the air, cool it on hot days, conserve heat at night, and act as excellent sound absorbers. Transpiration from the forests affects the relative humidity and precipitation in a place. Forests clean the environment by muffling noises, buffering strong winds and stopping dust and gases.

• The layer of leaves that fall around the tree prevents runoff and allows the water to percolate into the soil. Thus helping in ground water recharge.

• Dead plants decompose to form humus, organic matter that holds the water and provides nutrients to the soil.

- Through the process of photosynthesis, forests renew the oxygen supply in the atmosphere by absorbing atmospheric CO2 and moderating the greenhouse effect. As per the report published by Ministry of Environment and Forests during August 2009, the annual CO2 removal by India's forest and tree cover is enough to neutralize 11.25 % of India's total GHG emissions (CO2 equivalent) at 1994 levels. This is equivalent to offsetting 100% emissions from all energy in residential and transport sectors; or 40% of total emissions from the agriculture sector. Clearly, India's forest and tree cover is serving as a major mode of carbon mitigation for India and the world.

• Forest cover of an area plays an important role in amount of precipitation received by the area. Thus play an important role in maintaining water cycle of the area.

• Some species of trees have the ability to return nitrogen to the soil through root decomposition or fallen leaves. Such trees are planted to increase the nitrogen content of the soil.

• Forests absorb suspended particles in air thereby reducing pollution.

• Forests also helps in the process of soil formation by causing weathering of rock

• They play vital role in maintaining healthy watershed. Rivers originate in a forest area and carry the organic matter from forest to the downstream thus supporting a variety of fishes and aquatic animals. The richness of forest in upstream decides the biological value of the river ecosystem supported by it.

- It provides forest food which has great medicinal value and used by local people in respective season.

In performing all these functions, forest stabilizes the climate, maintains the ecological /environmental balance of the area and shape the landscape of the area.

Contribution to Economy

- It provides valuable items like timber, paper, fuel wood, bamboo, cane, food, fibers, essential oils.

- Forest plants provide hundreds of medicinal plants, spices, poisons, insecticides, soap substitutes like ritha and shikakai, tendu leaves used in bidi wrapping.

- Forests also provide fodder for cattle and other grazing animals. Leaves and twigs of some plants have high fodder value. It is useful fodder source during drought. In addition to this forests are also popular areas for relaxation & recreation and they add to the aesthetic value of the area.

Classification of Forests

Forests can be classified in different ways. The forest type depends upon the abiotic factors such as climate and soil characteristics of a region. Forests in India can be broadly divided into coniferous forests and broadleaved forests. They can also be classified according to the nature of their tree species- evergreen, deciduous, xerophytes or thorn trees, mangroves, etc. They can also be classified according to the most abundant species of trees, such as Sal or Teak forests. In many cases, a forest is named after the first three or four most abundant tree species.

Coniferous forests grow in the Himalayan mountain region, where the temperatures are low. These forests have tall stately trees with needle-like leaves and downward –sloping branches, so that the snow can slip off the branches.

Broad-leaved forests are of several types, such as evergreen forests, deciduous forests, thorn forests, and mangrove forests. Broad-leaved trees usually have large leaves of various shapes and are found in middle to lower latitude.

Evergreen forests grow in the high rainfall areas of the Western Ghats, North –eastern India and the Andaman and Nicobar Islands. These forests grow in areas where the monsoon period lasts for several months.

Deciduous forests are found in regions with a moderate amount of seasonal rainfall that lasts for only a few months. Most of the forests in which Teak trees grow are of this type. The deciduous trees shed their leaves during the winter and hot summer months.

Thorn forests are found in the semi-arid regions of India. The trees, which are sparsely distributed, are surrounded by open grassy areas.

Mangroves forests grow along the coast especially in the river deltas. These plants are uniquely adapted to be able to grow in a mix of saline and freshwater. They grow luxuriantly in muddy areas covered with silt that the rivers have brought down. The mangrove trees have breathing roots that emerge from the mud banks.

Forest Types in India

India has a diverse range of forests from the rainforest of Kerala in the south to the alpine pastures of Ladakh in the north, from the deserts of Rajasthan in the west to the evergreen forests in the north-east. Climate, soil type, topography, and elevation are the main factors that determine the type of forest. Forests are classified according to their nature and composition, the type of climate in which they thrive, and its relationship with the surrounding environment.

Champion & Seth system of classification (1968) provides an elaborate description of forest types of India in six major groups which are further divided into 16 type groups and finally into 200 types including subtypes and variations of forests. The 'forest type' may be defined as a unit of vegetation with distinctive physiognomy and structure. As per Champion & Seth, the determining factors of the forest types are climate, soil, vegetation and the past treatment (including biotic interference).

Table: Types of Forests of India

Moist tropical Forests Group 1- Wet evergreen Group 2 - Semi-evergreen Group 3 - Moist deciduous Group 4 - Littoral and swamp	**Montane sub tropical Forests** Group 8 - Broad leaved Group 9 - Pine Group 10 - Dry evergreen
Dry tropical Forests Group 5 - Dry deciduous Group 6 - Thorn Group 7 - Dry evergreen	**Montane temperate Forests** Group 11 - Wet Group 12 - Moist Group 13 - Dry
Group 14 - **Sub alpine forests**	**Alpine Forests** Group 15- Moist Group 16- Dry

Moist tropical forests

1. Wet evergreen

Wet evergreen forests are found in the south along the Western Ghats and the Nicobar and Andaman Islands and all along the north-eastern region. It is characterized by tall, straight evergreen trees that have a buttressed trunk or root on three sides like a tripod that helps to keep a tree upright during a storm. These trees often rise to a great height before they open out like a cauliflower. The more common trees that are found here are the jackfruit, betel nut palm, jamun, mango, and hollock. The trees in this forest form a tier pattern: shrubs cover the layer closer to the ground, followed by the short structured trees and then the tall variety. Beautiful fern of various colours and different varieties of orchids grow on the trunks of the trees.

2. Semi-evergreen

Semi-evergreen forests are found in the Western Ghats, Andaman and Nicobar Islands, and the Eastern Himalayas. Such forests have a mixture of the wet evergreen trees and the moist deciduous trees. The forest is dense and is filled with a large variety of trees of both types.

3. **Moist deciduous**

Moist deciduous forests are found throughout India except in the western and the north-western regions. The trees have broad trunks, are tall and have branching trunks and roots to hold them firmly to the ground. Some of the taller trees shed their leaves in the dry season. There is a layer of shorter trees and evergreen shrubs in the undergrowth. These forests are dominated by sal and teak, along with mango, bamboo, and rosewood.

4. **Littoral and swamp**

Littoral and swamp forests are found along the Andaman and Nicobar Islands and the delta area of the Ganga and the Brahmaputra. It consists mainly of whistling pines, mangrove dates, palms, and bulletwood. They have roots that consist of soft tissue so that the plant can breathe in the water.

Dry Tropical Forests

1. **Dry deciduous forest**

Dry deciduous forests are found throughout the northern part of the country except in the NorthEast. It is also found in Madhya Pradesh, Gujarat, Andhra Pradesh, Karnataka, and Tamil Nadu. The canopy of the trees does not normally exceed 25 metres. The common trees are the sal, a variety of acacia, and bamboo.

2. **Thorn**

This type is found in areas with black soil: North, West, Central, and South India. The trees do not grow beyond 10 metres. Spurge, caper, and cactus are typical of this region.

3. **Dry Evergreen**

Dry evergreens are found along the Andhra Pradesh and Karnataka coast. It has mainly hard leaved evergreen trees with fragrant flowers, along with a few deciduous trees.

Montane Sub-Tropical Forests

1. **Broad-leaved forests**

Broad-leaved forests are found in the Eastern Himalayas and the Western Ghats, along the Silent Valley. There is a marked difference in the form of the vegetation in the two areas. In the Silent Valley, the poonspar, cinnamon, rhododendron, and fragrant grass are predominant. In the Eastern Himalayas, the flora has been badly affected by the shifting cultivation and forest fires. These wet forests consist mainly of evergreen trees with a sprinkling of deciduous here and there. There are oak, alder, chestnut, birch, and cherry trees. There are a large variety of orchids, bamboo and creepers.

2. **Pine**

Pine forests are found in the steep dry slopes of the Shivalik Hills, Western and Central Himalayas, Khasi, Naga, and Manipur Hills. The trees predominantly found in these areas are the chir, oak, rhododendron, and pine. In the lower regions sal, sandan, amla, and laburnum are found.

3. **Dry Evergreen**

Dry evergreen forests normally have a prolonged hot and dry season and a cold winter. It generally has evergreen trees with shining leaves that have a varnished look. Some of the more common ones are the pomegranate, olive, and oleander. These forests are found in the Shivalik Hills and foothills of the Himalayas up to a height of 1000 metres.

Montane Temperate Forests

1. **Wet**

Wet montane temperate forests occur in the North and the South. In the North, it is found in the region to the east of Nepal into Arunachal Pradesh, at a height of 1800–3000 metres, receiving a minimum rainfall of 2000 mm. In the South, it is found in parts of the Niligiri Hills, the higher reaches of Kerala. The forests in the northern region are denser than in the South. This is because over time the original trees have been replaced by fast-growing varieties such as the eucalyptus. Rhododendrons and a variety of ground flora can be found here. In the North, there are three layers of forests: the higher layer has mainly coniferous, the middle layer has deciduous trees such as the oak and the lowest layer is covered by rhododendron and champa.

2. **Moist**

This type spreads from the Western Himalayas to the Eastern Himalayas. The trees found in the western section are broad-leaved oak, brown oak, walnut, rhododendron, etc. In the Eastern Himalayas, the rainfall is much heavier and therefore the vegetation is also more lush and dense. 6 There are a large variety of broad-leaved trees, ferns, and bamboo. Coniferous trees are also found here, some of the varieties being different from the ones found in the South.

3. **Dry**

This type is found mainly in Lahul, Kinnaur, Sikkim, and other parts of the Himalayas. There are predominantly coniferous trees that are not too tall, along with broad-leaved trees such as the oak, maple, and ash. At higher elevation, fir, juniper, deodar, and chilgoza can be found.

Sub Alpine Forests

Sub alpine forests extend from Kashmir to Arunachal Pradesh between 2900 to 3500 metres. In the Western Himalayas, the vegetation consists mainly of juniper, rhododendron, willow, and black currant. In the eastern parts, red fir, black juniper, birch, and larch are the common trees. Due to heavy rainfall and high humidity the timberline in this part is higher than that in the West. Rhododendron of many species covers the hills in these parts.

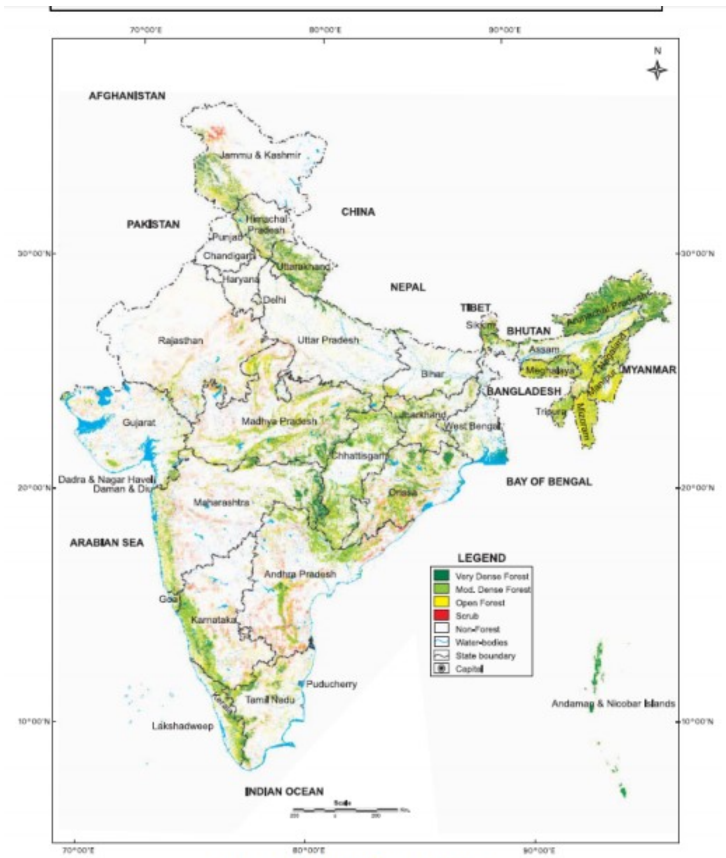
Alpine Forests

1. **Moist**

Moist alpines are found all along the Himalayas and on the higher hills near the Myanmar border. It has a low scrub, dense evergreen forest, consisting mainly of rhododendron and birch. Mosses and ferns cover the ground in patches. This region receives heavy snowfall.

2. **Dry**

Dry alpines are found from about 3000 metres to about 4900 metres. Dwarf plants predominate, mainly the black juniper, the drooping juniper, honeysuckle, and willow.



Forest Cover Map of India

Source: India State of Forest Report, 2009, Forest Survey of India.

Threats to Forests

As the rate of development is increasing, it is putting pressure on all the natural resources around us. Forests are also getting depleted at a fast rate all over the world. Over use and irrational use, technological and industrial growth, population growth and increasing consumption levels are major factors causing depletion of forest resources. Some other factors are mining, submergence due to big dams, shifting cultivation, use of forestlands for rehabilitation, agriculture, transport and tourism. All these activities are causing qualitative as well as quantitative depletion of forests.

The forest wood is used up for construction, furniture, deriving energy (coal and firewood) and thus the increasing demand for timber, energy, paper and paper products has led to massive destruction of forests. When forest is cut, it is not just the trees that go but the entire ecosystem is lost which is invaluable. The full grown forests, existing since thousands of years can't be replaced by plantations. As forests grow very slowly, people cannot use more resources than they can produce during a growing season. If timber is felled beyond a certain limit, the forest cannot regenerate. The gaps in the forest change the habitat quality for its animal, and the more sensitive species cannot survive under these changed conditions.

Over-utilizing forest resources is an unsustainable way of using our limited forest resources. As the forest resources are exploited, the forest canopy is opened up, the ecosystem is degraded, and its wildlife is seriously threatened. Increasing tourism activities are also causing destruction to forest ecosystem. when the frequency of visitors and tourists becomes excessive, problems of soil erosion occur along and adjacent to footpaths.

Wildlife is also disturbed, plants and saplings are trampled, and the waste is dumped at these places which disturbs entire ecosystem. Forest fire is also an important threat to forests, which destroy large areas of forest every year all over the world. It has detrimental impact on forest, wildlife and people living around. Along with the loss of biodiversity, it results in increased air pollution, migration of animals to different areas. It directly affects the livelihood of people dependant on forest resources. Natural disasters like Tsunami, earthquake destroys large forests areas in a short span of time. All these natural as well as man-made factors cause destruction of forests at a large scale.

Forest Decline

According to the World Resources Institute, the world has lost about half of its forest cover. Despite a number of initiatives to stop forest decline, the world continues to lose some 15 million hectares of forests every year. Deforestation over the period 1980-1990 reached 8.2% of total forest area in Asia, 6.1% in Latin America and 4.8% in Africa. Most modern deforestation takes place in developing countries, particularly in tropical areas. The process generates large amounts of carbon dioxide – equivalent to 20% of global emissions from fossil fuels, making deforestation the second most important contributor to global warming – and results in annual degradation of some 12 million hectares of fertile land and loss of thousands of species (estimates range between 8,000 and 28,000 per year). Deforestation and forest degradation directly threaten as many as 400 million people – including 50 million forest indigenous people – who depend on forests for subsistence. Forest decline, resulting from the enormous human ability to alter large forest ecosystems is the source of intense conflicts between rural populations, governments, commercial interests and, increasingly, sections of the public at large.

Forest decline: What is it?

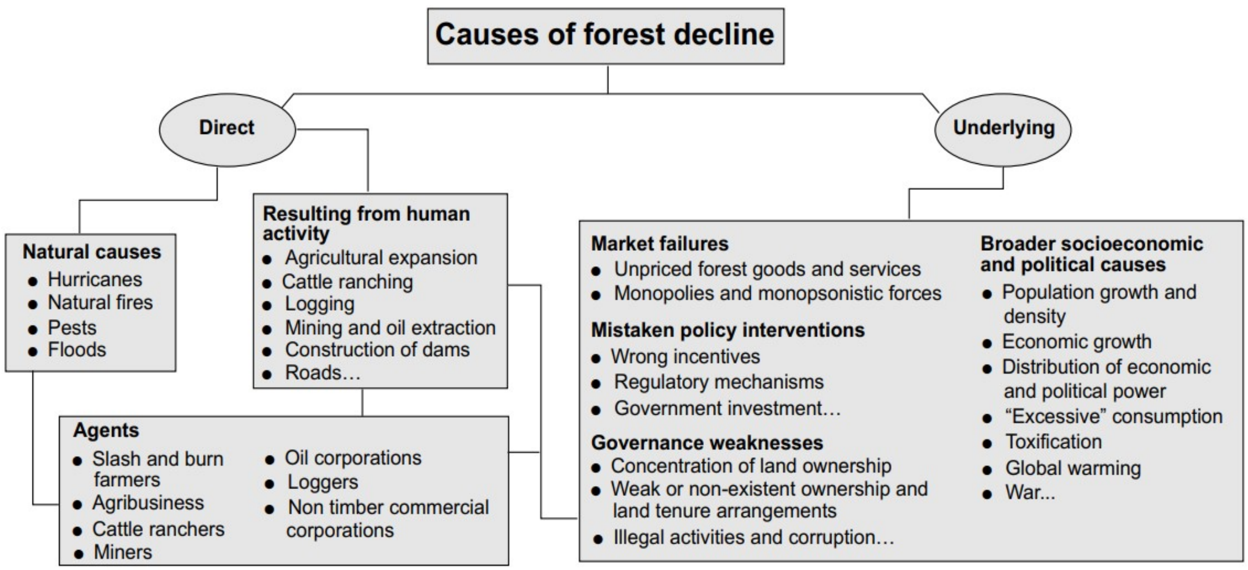
Forest decline here is interpreted as deforestation, forest degradation or a combination of both. These terms are not precise. The Food and Agriculture Organization of the United Nations defines deforestation as the "sum of all transitions from natural forest classes (continuous and fragmented) to all other classes" (FAO 1997). The loss of forest cover attributed to these transitions must occur over less than 10% of the crown cover for the phenomenon to qualify as deforestation.

It is not clear whether this refers to substantial areas or to, say, a particular hectare. What is the area that can be considered as under "fragmented" transition? Should such transition involve a minimum of, for example, 10 hectares to qualify as deforestation? FAO suggests that a minimum of 0.5 hectare is needed to qualify as forest and therefore loss of crown cover to less than 10% for at least this area would presumably qualify as deforestation (see, for example, FAO 1994). This, however, is not universally accepted. Operationally, this level of precision (0.5 ha) may be unattainable, particularly in developing countries.

In addition to deforestation, forest degradation is an issue. According to FAO, changes within a forest class, for example from closed to open forest, which negatively affect the stand or, in particular, lower its production capacity, constitute forest degradation. Thus, forest degradation implies a major loss of forest productive capacity, even where there is little deforestation as such. FAO (1998) states:

Forest degradation takes different forms, particularly in open forest formations, deriving mainly from human activities such as overgrazing, over-exploitation (for firewood or timber), repeated fires, or due to attack by insects, diseases, plant parasites or other natural causes such as cyclones. In most cases, degradation does not show as a decrease in the area of woody vegetation but rather as a gradual reduction of biomass, changes in species composition and soil degradation. Unsustainable logging practices can contribute to degradation if extraction of mature trees is not accompanied with their regeneration or if the use of heavy machinery causes soil compaction or loss of productive forest area.

Causes of Forest Decline



Source: Contreras-Hermosilla, 2000: 5

Mistakenly Policy Interventions

Policy interventions may create obstacles to the sustainable management of forests. These actions are not always purposely biased against forests but often this is their unintended result. In many cases they constitute a failure to address market failures. As expressed by Repetto (1993): Governments, many of which are committed in principle to conservation and wise resource use, are aggravating the loss of the forests under their stewardship through mistaken policies. Such policies, by and large, were adopted for worthy objectives: industrial or agricultural growth, regional development, job creation, or poverty alleviation. But such objectives typically have not been realised or have been attained only at excessive cost.

Government-sponsored road construction and the onset of deforestation are frequently easily observable circumstances in various tropical countries. For example, in the Brazilian state of Pará, deforestation following road construction increased from 0.6% to 17.3% of the state's area between 1972 and 1985. Although in theory some of the negative impacts of roads policies could be mitigated, in practice governments do not have the inclination to do so for political reasons or simply because they cannot, given the limited administrative resources at their command.

Subsidy Policies

As with other underlying causes of forest decline, the cause-effect link between faulty concession contracts and deforestation and forest degradation is not always clear. Some governments establish subsidised timber concessions to open lands to economic opportunity and to provide means of livelihood for impoverished migrants. Governments occasionally justify subsidies embodied in timber concessions on grounds of employment creation, construction of infrastructure by logging companies and the push to promote local development. Generally, however, these are inefficient methods of achieving local development and beneficial impacts are seldom sustainable.

Whether agricultural subsidies and the associated possibility to increase agricultural profitability will lead to either intensification (and reduced pressure on forests, as in the case of the green revolution) or extensification depends on the technologies adopted, the economic conditions facing farmers, the availability of different types of lands and the nature of subsidies.

Table: Deforestation by Sub-regions in between 1990-2010

Region/subregion	1990-2000		2000-2010	
	1 000 ha/year	%	1 000 ha/year	%
Eastern and Southern Africa	-1841	-0.62	-1839	-0.66
Northern Africa	-590	-0.72	-41	-0.05
Western and Central Africa	-1637	-0.46	-1535	-0.46
Total Africa	**-4067**	**-0.56**	**-3414**	**-0.49**
East Asia	1762	0.81	2781	1.16
South and Southeast Asia	-2428	-0.77	-677	-0.23
Western and Central Asia	72	0.17	131	0.31
Total Asia	**-595**	**-0.10**	**2235**	**0.39**
Russian Federation (RF)	32	n.s.	-18	n.s.
Europe excluding RF	845	0.46	694	0.36
Total Europe	**877**	**0.09**	**676**	**0.07**
Caribbean	53	0.87	50	0.75
Central America	-374	-1.56	-248	-1.19
North America	32	n.s.	188	0.03
Total North and Central America	**-289**	**-0.04**	**-10**	**0.00**
Total Oceania	**-41**	**-0.02**	**-700**	**-0.36**
Total South America	**-4213**	**-0.45**	**-3997**	**-0.45**
World	**-8327**	**-0.20**	**-5211**	**-0.13**

Source: Anon., 2010

- ❖ **Direct Causes**

Expansion of Farming Land

About 60 per cent of the clearing of tropical moist forests is for agricultural settlement (Myers, 1994; Anon., 1991) with logging and other reasons like roads, urbanization and Fuel-wood accounting for the rest (Anon., 1994b). Tropical forests are one of the last frontiers in the search for subsistence land for the most vulnerable people worldwide (Myers, 1992). Millions of people live on the tropical forest with less than a dollar a day where a third of a billion are estimated to be foreign settlers. However, as the land degrades people are forced to migrate, exploring new forest frontiers increasing deforestation (Wilkie et al., 2000; Amor, 2008; Amor and Pfaff, 2008).

Deforestation is proxied by the expansion of agricultural land. This is because agricultural land expansion is generally viewed as the main source of deforestation contributing around 60 per cent of total tropical deforestation. Shifting agriculture also called slash and burn agriculture is the clearing of forested land for raising or growing the crops until the soil is exhausted of nutrients and/or the site is overtaken by weeds and then moving on to clear more forest. It is been often reported as the main agent of deforestation. Smallholder production in deforestation and the growing number of such producers notably shifting cultivators were the main cause of deforestation (Anon., 1990b; c; Dick, 1991; Anon., 1992a; b; Barbier et al., 1993; Ascher, 1993; Dove, 1993; 1996; Dauvergne, 1994; Porter, 1994; Thiele, 1994; Anon., 1994c; Angelsen, 1995; Ross, 1996). Mostly all reports indicate shifting agriculture as responsible for about one half of tropical deforestation and some put it up to two-thirds. Shifting agriculture was greatest in Asia (about 30 per cent) but only about 15 per cent over the whole tropical world. It appears that the proportion of direct conversion of forest to agriculture is increasing and the proportion of shifting agriculture is decreasing with time.

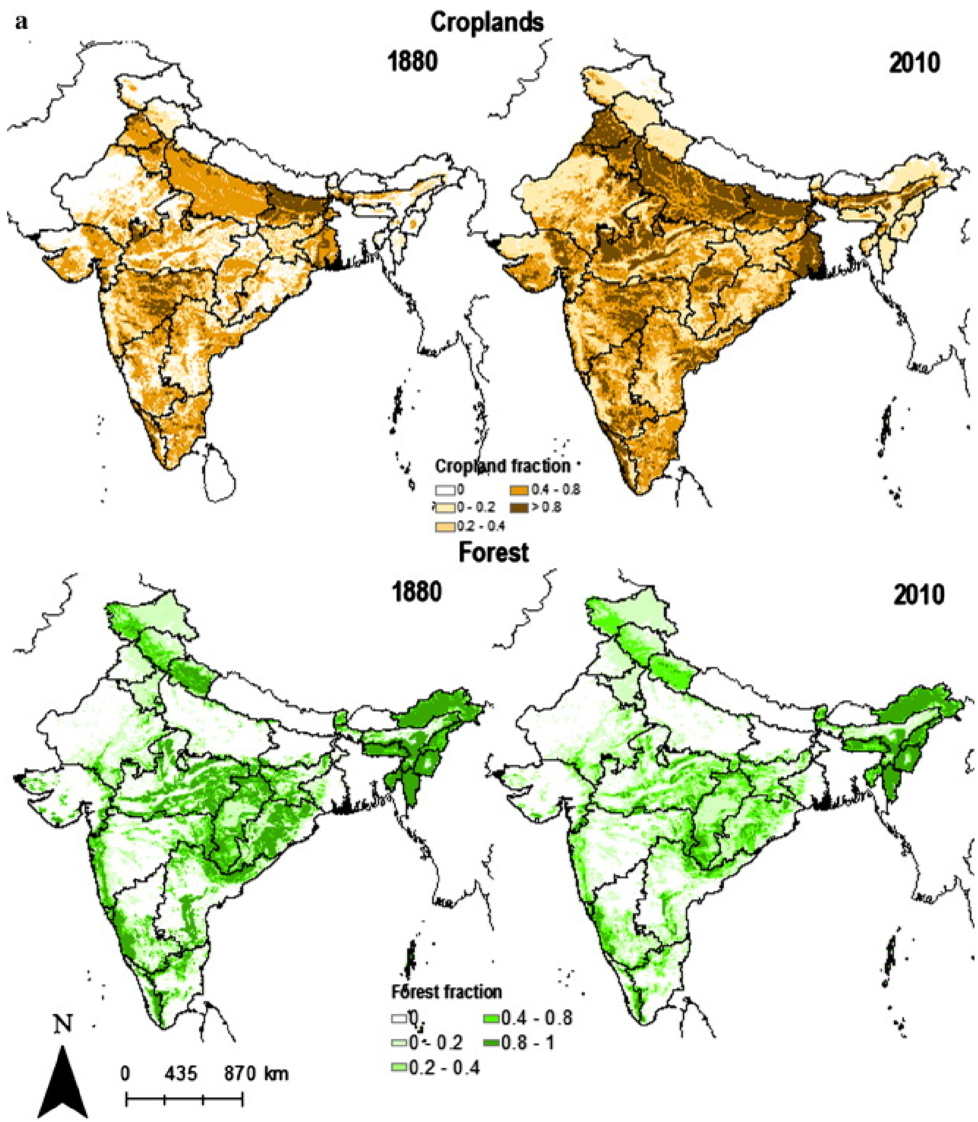


Figure: Deforestation and Expansion of Crop Lands, 1980-2010

Source: sciencedirect.com

Forest and Other Plantations

Plantations are a positive benefit and should assist in reducing the rate of deforestation. The fact that plantations remove the timber pressure on natural forests does not translate eventually into less, but rather into more deforestation. Indeed, it is feared that agricultural expansion which is the main cause of deforestation in the tropics might replace forestry in the remaining natural forests (Anon., 2002; Cossalter and Pye-Smith, 2003; Anon., 2005). The impact of timber plantations could thus turn out to be quite detrimental to tropical forest ecosystems (Kartodihardjo and Supriono, 2000). Tree crops and rubber in particular plays a more important role in deforestation in Indonesia than subsistence-oriented shifting cultivation (Chomitz and Griffiths, 1996). Unfortunately about one-half of the plantations in the tropics are established on native forest cleared for the purpose. Moreover plantations can promote deforestation by constructing roads that improve access of the shifting cultivators and others to the forest frontier.

Logging and Fuel Wood

Logging does not necessarily cause deforestation. However, logging can seriously degrade forests (Putz et al., 2001). Logging in Southeast Asia is more intensive and can be quite destructive. However, logging provides access roads to follow-on settlers and log scales can help finance the cost of clearing remaining trees and preparing land for planting of crops or pasture. Logging thus catalyzes deforestation (Chomitz et al., 2007). Fuelwood gathering is often concentrated in tropical dry forests and degraded forest areas (Repetto, 1988; 1990; Rowe et al., 1992; Anon., 1994a). Fuelwood is not usually the major cause of deforestation in the humid tropics although it can be in some populated regions with reduced forest area such as in the Philippines, Thailand and parts of Central America. Fuelwood gathering was considered to be the main cause of deforestation and forest degradation in El Salvador (Repetto, 1990). In the drier areas of tropics, Fuelwood gathering can be a major cause of deforestation and degradation.

Overgrazing

Overgrazing is more common in drier areas of the tropics where pastures degraded by overgrazing are subject to soil erosion. Stripping trees to provide fodder for grazing animals can also be a problem in some dry areas of the tropics but is probably not a major cause of deforestation. Clear cutting and overgrazing have turned large areas of Qinghai province in China into a desert. Overgrazing are causing large areas of grasslands north of Beijing and in Inner Mongolia and Qinghai province to turn into a desert. One man who lived in a village on the eastern edge of the Qinghai-Tibet plateau that was being swallowed up by sand told the New York Times, "The pasture here used to be so green and rich. But now the grass is disappearing and the sand is coming." Huge flocks of sheep and goats strip the land of vegetation. In Xillinggol Prefecture in Inner Mongolia, for example, the livestock population increased from 2 million in 1977 to 18 million in 2000, turning one third of the grassland area to desert. Unless something is done the entire prefecture could be uninhabitable by 2020.

Overgrazing is exacerbated by sociological phenomena called "the tragedy of the common." People share land but raises animals for themselves and try to enrich them by rising as many as they can. This leads to more animals than the land can support. Grassland in Qinghai that can support 3.7 million sheep had 5.5 million sheep in 1997. Animals remove the vegetation and winds finished the job by blowing away the top soil, transforming grasslands into desert. When a herder was asked why he was grazing goats next to a sign that said "Protect vegetation, no grazing," he said, "The lands are too infertile to grow crops—herding is the only way for us to survive." (Hays, 2008 web page).

Fires

Fires are a major tool used in clearing the forest for shifting and permanent agriculture and for developing pastures. Fire is a good servant but has a poor master. Fire used responsibly can be a valuable tool in agricultural and forest management but if abused it can be a significant cause of deforestation (Repetto, 1988; Rowe et al., 1992). Based on the data available from 118 countries representing 65 per cent of the global forest area, an average of 19.8 million hectares or one per cent of all forests were reported to be significantly affected each year by forest fires (Anon., 2010). Deforestation due to road pavements in Brazil had also lead to higher incidences of forest fires (Carvalho et al., 2001; Nepstad et al., 2001).

Mining

Mining is very intensive and very destructive (Mather, 1991; Sands, 2005). The area of land involved is quite small and it is not seen as a major cause of primary deforestation. Mining is a lucrative activity promoting development booms which may attract population growth with consequent deforestation. The deforestation rate due to mining activities in Guyana from 2000 to 2008 increased 2.77 times according to an assessment by the World Wildlife Fund-Guianas (Staff, 2010). Similarly, in the Philippines, mining, along with logging, has been among the forces behind the country's loss of forest cover: from 17 million hectares in 1934 to just three million in 2003 or an 82 per cent decline (Docena, 2010). Nearly 2,000 hectares of tropical forest in the Municipality of Coahuayana in the State of Michoacán (south-western Mexico) will completely be destroyed by mining iron minerals planned by the Italo-Argentine mining company TERNIUM (Anonymous, 2008). Similarly, Nyamagari hills in Orissa India currently threatened by Vedanta Aluminum Corporation's plan to start bauxite mining will destroy 750 hectares of reserved forest (Griffiths and Hirvelä, 2008).

Massive and unchecked mining of coal, iron ore and bauxite in Jharkhand, India has caused large scale deforestation and created a huge water scarcity (Anon., 2011b). In return for US$3.8 billion of investment, the agreements between the State government of Jharkhand, India and mining companies, there will be a massive land acquisition which will deforest no less than 57,000 hectares of forest and displace 9,615 families, many of them located in legally protected Scheduled Areas set aside for indigenous peoples in the State (Mullick and Griffiths. 2007). Moreover, Roads constructed to support the mining operations will open up the area to shifting agriculturists, permanent farmers, ranchers, land speculators and infrastructure developers.

For instance the core of Brazil's Amazon development strategy were infra-structure development projects such as roads providing access to frontier regions, mining area and large hydroelectric reservoirs (Mahar, 1988; Fearnside and Barbosa, 1996; Carvalho et al., 2002, 2004). The construction of roads, railways, bridges, and airports opens up the land to development and brings increasing numbers of peoples to the forest frontier. If wood is used as fuel in mining operations and it is sources from plantations established for the purpose, it can cause serious deforestation in the region. On the other hand, mining can be labour intensive and take labour away from clearing forest.

Urbanization/industrialization and Infra-structure

Expanding cities and towns require land to establish the infrastructures necessary to support growing population which is done by clearing the forests (Mather, 1991; Sands, 2005). Tropical forests are a major target of infra-structure developments for oil exploitation, logging concessions or hydropower dam construction which inevitably conveys the expansion of the road network and the construction of roads in pristine areas (Kaimowitz and Angelsen, 1998). The construction of roads, railways, bridges, and airports opens up the land to development and brings increasing numbers of people to the forest frontier. Whether supported or not by the governmental programmes, these settlers have usually colonized the forest by using logging trails or new roads to accessthe forest for subsistence land (Wilkie et al., 2000; Amor, 2008; Amor and Pfaff, 2008). (Wilkie et al., 2000; Amor, 2008; Amor and Pfaff, 2008). The development of these infrastructure projects are of worldwide concern, since tropical forest clearing accounts for roughly 20 per cent of anthropogenic carbon emissions destroying globally significant carbon sinks (Anon., 2001c) and around 21 per cent of tropical forests have been lost worldwide since 1980 (Bawa et al., 2004).

Air pollution

Air pollution is associated with degradation of some European and North American forests. The syndrome is called "Waldsterben" or forest death. In 1982, eight per cent of all West German trees exhibited damage that rose to about 52 per cent by 1987 (Raloff, 1989) and half of the trees reported dying of Waldsterben in the Alps (Lean, 1990). High elevation forests show the earliest damage including forests in the north-east and central United States.

Wars and Role of the Military

It is well established that military operations caused deforestation during the Vietnam War and elsewhere (Mather, 1991; Sands, 2005). More recently, linkages have been documented between the civil war in Myanmar and the timber trade between Myanmar and Thailand. Myanmar regime sells timber to the Thais to finance its civil war against the Karen hill tribe. Forest destruction in El Salvador has resulted from war. Apart from military involvements in wars, the role of military in deforestation has been documented in Southeast Asia and South America (Mather, 1991; Sands, 2005). The authors also observed that role of powerful military in Brazilian politics are a major cause of Amazonian forest destruction.

Tourism

National parks and sanctuaries beyond doubt protect the forests, but uncautioned and improper opening of these areas to the public for tourism is damaging. Unfortunately, the national governments of tropical and sub-tropical countries adopt tourism for easy way of making money sacrificing the stringent management strategies. Further, many companies and resorts who advertise themselves as eco-tourist establishments are in fact exploiting the forests for profit. In Cape Tribulation, Australia, for example, the rain forest is being threatened by excessive tourism (Colchester and Lohmann, 1993). Similarly, in the Terai Duars of eastern India foothill Himalaya, eco-tourism is encouraged and we fear this is being done without developing adequate management plans.

For instance, the Chilapatta Reserve Forest in this area is opened for eco-tourism for its ancient ruins deep in the forest and a tree species *Myristica longifolia* that exudes a blood like sap when injured. The site has become a popular eco-tourist destination because of the ruins and for this blood exuding tree. In the whole forest only eight individuals were found but two of the trees in the near vicinity of the ruins completely dried away due to repeated injuries caused to the plants by the curious tourists (Shukla, 2010). In fact, in the name of eco-tourism, infra-structure development is taking place mostly be the private players in these wilderness areas which are further detrimental in terms of attracting peoples other than tourists also, causing deforestation especially deep in the forest.

- ❖ **Indirect Causes**

The World Rainforest Movement's 'Emergency Call to Action for the forests and their Peoples' asserts that "deforestation is the inevitable result of the current social and economic policies being carried out in the name of development" (Anon., 1990d). It is in the name of development that irrational and unscrupulous logging, cash crops, cattle ranching, large dams, colonisation schemes, the dispossession of peasants and indigenous peoples and promotion of tourism is carried out. Harrison Ngau, an indigenous tribesman from Sarawak, Malaysia and winner of the Goldman Environment Award in 1990 puts the cause of tropical deforestation like this, "the roots of the problem of deforestation and waste of resources are located in the industrialized countries where most of our resources such as tropical timber end up. The rich nations with one quarter of the world's population consume four fifth of the world's resources. It is the throw away culture of the industrialized countries now advertised in and forced on to the Third World countries that is leading to the throwing away of the world. Such so-called progress leads to destruction and despair" (Anon., 1990d)! Such a development leads to overconsumption which is the basic underlying cause of deforestation.

Colonialism

Erstwhile colonies of the colonial powers like Britain, France, Spain or Portugal are now the Third World Countries or the developing nations mostly have the tropical rainforests except Australia and Hawaii were exploited for their natural resources and their indigenous people's rights destroyed by the colonial powers. All these countries have indigenous populations who had their own system of land management and/or ownership in place for thousands of years before the intervention of colonists from rich industrialized nations. Colonialism turned previously self-sufficient economies into zones of agriculture export production. This process continues even today in different form of exploitation and the situation is worsening (Colchester and Lohmann, 1993).

Exploitation by Industrialized Countries

Wealthy countries or the erstwhile colonial powers having deficit of their own natural resources are mainly sustaining on the resources of the financially poorer countries those are generally natural resource rich. Twenty per cent of the world's population is using 80 per cent of the world's resources. Unfortunately also the governments of these poor resource rich countries had generally adopted the same growth-syndrome as their western neighbours or their erstwhile colonial master giving emphasis on maximizing exports, revenues and exploiting their rich natural resources unsustainably for short-term gains. Moreover, corruption in government, the military and economic powers is well known. The problem is further worsened by the low price of the most Third World exports being realized in the international market (Colchester and Lohmann, 1993).

The debt burden

Pursuing the guided development agenda, the financially poorer countries are on a heavy international debt and now feeling the urgency of repaying these huge debts due to escalating interest rates. Such a situation compels these debt ridden poorer countries toexploit their rich natural resources including their forests partly to earn foreign exchange for servicing their debts. For instance, construction of roads for logging operations in some South-east Asian countries was funded by Japanese aid which allowed the Japanese timber companies to exploit the forests of these countries. Understandably, these timber companies profitably exploited the forests while the South-east Asian countries were left owing Japan money for construction of their roads (Colchester and Lohmann, 1993).

Overpopulation and Poverty

The role of population in deforestation is a contentious issue (Mather, 1991; Colchester and Lohmann, 1993; Cropper and Griffiths, 1994; Ehrhardt-Martinez, 1998; Sands, 2005). The impact of population density on deforestation has been a subject of controversy. Poverty and overpopulation are believed to be the main causes of forest loss according to the international agencies such as FAO and intergovernmental bodies. It is generally believed by these organizations that they can solve the problem by encouraging development and trying to reduce population growth. Conversely, the World Rainforest Movement and many other NGOs hold unrestrained development and the excessive consumption habits of rich industrialized countries directly responsible for most forest loss.

However there is good evidence that rapid population growth is a major indirect and over-arching cause of deforestation. More people require more food and space which requires more land for agriculture and habitation. This in turn results in more clearing of forests. Arguably increasing population is the biggest challenge of all to achieve sustainable management of human life support systems and controlling population growth is perhaps the best single thing that can be done to promote sustainability. Overpopulation is not a problem exclusive to Third World countries. An individual in an industrialized country is likely to consume in the order of sixty times as much of the world's resources as a person in a poor country. The growing population in rich industrialized nations are therefore responsible for much of the exploitation of the earth and there is a clear link between the overconsumption in rich countries and deforestation in the tropics (Colchester and Lohmann, 1993).

Corruption and Political Cause

The FAO identified forest crime and corruption as one of the main causes of deforestation in its 2001 report and warned that immediate attention has to be given to illegal activities and corruption in the world's forests in many countries (Anon., 2001b). Illegal forest practices may include the approval of illegal contracts with private enterprises by forestry officers, illegal sale of harvesting permits, under-declaring volumes cut in public forest, under-pricing of wood in concessions, harvesting of protected trees by commercial corporations, smuggling of forest products across borders and allowing illegal logging, processing forest raw materials without a license (Contreras-Hermosilla, 2000; 2001).

Effects of Deforestation

Climate Change

It is essential to distinguish between microclimates, regional climate and global climate while assessing the effects of forest on climate (Gupta et al., 2005) especially the effect of tropical deforestation on climate (Dickinson, 1981). Deforestation can change the global change of energy not only through the micrometeorological processes but also by increasing the concentration of carbon dioxide in the atmosphere (Pinker, 1980) because carbon dioxide absorbs thermal infrared radiation in the atmosphere. Moreover deforestation can lead to increase in the albedo of the land surface and hence affects the radiation budget of the region (Charney, 1975; Rowntree, 1988; Gupta et al., 2005). Deforestation affects wind flows, water vapour flows and absorption of solar energy thus clearly influencing local and global climate (Chomitz et al., 2007). Deforestation on lowland plains moves cloud formation and rainfall to higher elevations (Lawton et al., 2001).

Deforestation disrupts normal weather patterns creating hotter and drier weather thus increasing drought and desertification, crop failures, melting of the polar ice caps, coastal flooding and displacement of major vegetation regimes. In the dry forest zones, land degradation has become an increasingly serious problem resulting in extreme cases in desertification (Dregne, 1983). Desertification is the consequence of extremes in climatic variation and unsustainable land use practices including overcutting of forest cover (Anon., 1994b). Global warming or global change includes anthropogenically produced climatic and ecological problems such as recent apparent climatic temperature shifts and precipitation regimes in some areas, sea level rise, stratospheric ozone depletion, atmospheric pollution and forest decline. Tropical forests are shrinking at a rate of about five per cent per decade as forests are logged and cleared to supply local, regional, national and global markets for wood products, cattle, agricultural produce and biofuels (Anon., 2007; 2010).

One of the most important ramifications of deforestation is its effect on the global atmosphere. Deforestation contributes to global warming which occurs from increased atmospheric concentrations of greenhouse gases (GHG) leading to net increase in the global mean temperature as the forests are primary terrestrial sink of carbon. Thus deforestation disrupts the global carbon cycle increasing the concentration of atmospheric carbon dioxide. Tropical deforestation is responsible for the emission of roughly two billion tonnes of carbon (as CO_2) to the atmosphere per year (Houghton, 2005). Release of the carbon dioxide due to global deforestation is equivalent to an estimated 25 per cent of emissions from combustion of fossil fuels (Asdrasko, 1990).

Water and Soil Resources Loss and Flooding

Deforestation also disrupts the global water cycle (Bruijnzeel, 2004). With removal of part of the forest, the area cannot hold as much water creating a drier climate. Water resources affected by deforestation include drinking water, fisheries and aquatic habitats, flood/drought control, waterways and dams affected by siltation, less appealing waterrelated recreation, and damage to crops and irrigation systems from erosion and turbidity (Anon., 1994a; Bruijnzeel et al., 2005). Urban water protection is potentially one of the most important services that forest provides (Chomitz et al., 2007). Filtering and treating water is expensive. Forests can reduce the costs of doing so either actively by filtering runoff or passively by substituting for housing or farms that generate runoff (Dudley and Stolton, 2003).

Deforestation can also result into watersheds that are no longer able to sustain and regulate water flows from rivers and streams. Once they are gone, too much water can result into downstream flooding, many of which have caused disasters in many parts of the world. This downstream flow causes soil erosion thus also silting of water courses, lakes and dams. Deforestation increases flooding mainly for two reasons. First, with a smaller ‘tree fountain’ effect, soils are more likely to be fully saturated with water. The ‘sponge’ fills up earlier in wet season, causing additional precipitation to run off and increasing flood risk. Second, deforestation often results in soil compaction unable to absorb rain. Locally, this causes a faster response of stream flows to rainfall and thus potential flash flooding (Chomitz et al., 2007).

Moreover deforestation also decrease dry season flows. The long term effect of deforestation on the soil resource can be severe. Clearing the vegetative cover for slash and burn farming exposes the soil to the intensity of the tropical sun and torrential rains. Forest floors with their leaf litter and porous soils easily accommodate intense rainfall. The effects of deforestation on water availability, flash floods and dry season flows depend on what happens to these countervailing influences of infiltration and evapotranspiration- the sponge versus the fountain (Bruijnzeel, 2004). Deforestation and other land use changes have increased the proportion of the basin subject to erosion and so over the long run have contributed to siltation. Heavy siltation has raised the river bed increasing the risk of flooding especially in Yangtze river basin in China, the major river basins of humid tropics in East Asia and the Amazonian basin (Yin and Li, 2001; Bruijnzeel, 2004; Aylward, 2005, Bruijnzeel et al., 2005; van Noordwijk et al., 2006).

Decreased Biodiversity, Habitat Loss and Conflicts

Forests especially those in the tropics serve as storehouses of biodiversity and consequently deforestation, fragmentation and degradation destroys the biodiversity as a whole and habitat for migratory species including the endangered ones, some of which have still to be catalogued. Tropical forests support about two thirds of all known species and contain 65 per cent of the world’s 10, 000 endangered species (Myers and Mittermeier, 2000). Retaining the biodiversity of the forested areas is like retaining a form of capital, until more research can establish the relative importance of various plants and animal species (Anon., 1994a).

According to the World Health Organization, about 80 per cent of the world’s population relies for primary health care at least partially on traditional medicine. The biodiversity loss and associated large changes in forest cover could trigger abrupt, irreversible and harmful changes. These include regional climate change including feedback effects that could theoretically shift rainforests to savannas and the emergence of new pathogens as the growing trade in bushmeat increases contact between humans and animals (Anon., 2005).

Another negative effect of deforestation is increasing incidents of human-animal conflicts hitting hard the success of conservation in a way alienating the people's participation in conservation. Elephant habitat located at northern West Bengal in India is part of the Eastern Himalaya Biodiversity Hotspot which is characterized by a high degree of fragmentation. The heavy fragmentation of this habitat has resulted into an intense human-elephant conflict causing not only in loss of agricultural crops but also human and elephant lives. Mortality of about 50 persons and 20 elephants was reported due to these severe human-elephant conflicts from this hotspot area annually (Sukumar et al., 2003; Mangave, 2004).

Economic Losses

The tropical forests destroyed each year amounts to a loss in forest capital valued at US $ 45 billion (Hansen, 1997). By destroying the forests, all potential future revenues and future employment that could be derived from their sustainable management for timber and non-timber products disappear.

Social Consequences

Deforestation, in other words, is an expression of social injustice (Colchester and Lohmann, 1993). The social consequences of deforestation are many, often with devastating long-term impacts. For indigenous communities, the arrival of civilization usually means the destruction/change of their traditional life-style and the breakdown of their social institutions mostly with their displacement from their ancestral area. The intrusion of outsiders destroys traditional life styles, customs and religious beliefs which intensifies with infra-structure development like construction of roads which results into frontier expansion often with social and land conflicts (Schmink and Wood, 1992).

The most immediate social impact of deforestation occurs at the local level with the loss of ecological services provided by the forests. Forests afford humans valuable services such as erosion prevention, flood control, water treatment, fisheries protection and pollination functions that are particularly important to the world's poorest people who rely on natural resources for their everyday survival. By destroying the forests we risk our own quality of life, gamble with the stability of climate and local weather, threaten the existence of other species and undermine the valuable services provided by biological diversity.

Managing Forests

Conservation, protection & overall management of forests are important. Entities involved in it & their role can be summarized as:

A. Government
B. Community and community organizations
C. Individuals and Private Bodies

A) Government

Making policy & laws for conservation, protection & overall management of resources is one of the important aspects of government initiatives. India's first Forest Policy was enunciated in 1952. Between 1952 and 1988, the extent of deforestation was so great that it became essential to 22 formulate a new policy on forests and their utilization. The earlier forest policies had focused only on revenue generation. In the 1980s it became clear that forests must be protected for their other functions such as the maintenance of soil and water regimes centered on ecological concerns. Thus the role of India's forests in the national economy and ecology has been reemphasized in the National Forest Policy, 1988, which focuses on ensuring environmental stability, restoring the ecological balance and preserving the forests.

The policy aims at increasing the forest and tree cover to 33 % of the country's land area. It also provided for the use of goods and services of the forest for its local inhabitants. Various laws like Forest Conservation Act, Wildlife Protection Act have been prepared and are being implemented by the government. The Forest Conservation Act of 1980 was enacted to control deforestation. In 1988, this act was amended to facilitate stricter conservation measures. Under the Wildlife Protection Act, 1972, different animals are categorized under specific schedules and there are restrictions on their hunting, as per this act specific ban is imposed on hunting of all Schedule 1 animals like tigers, leopards etc.

Another important government initiative is declaration of protected areas for conservation of wildlife. Sanctuaries and National Parks play a very important role in conservation of plants and animal gene pool. Several of these are present in catchment areas of major rivers, and so they are very important for water conservation and conservation of entire ecosystem. The government also implements many schemes for conservation of forests and for their sustainable management.

Protection & Management of Natural / Man-made Forests

• Natural and man-made forests / plantations on Hill slopes in catchment area of rivers and lakes will be protected to maintain ecological balance and environmental stability.

• Competent system will be developed to take strict action against the people / their actions that cause destruction to forests. Necessary changes will be done in Forest related laws to frighten the people about punishment given in forest related crime. A time limit will be set to prevent the uncontrolled grazing in forest areas. Modern technologies will be used for forest fire prevention.

• Forest guards will be well-equipped with modern communication systems (wireless machine, mobiles, and vehicles) and weapons to protect the natural and man-made forests from criminal gangs / poachers harming forests. Forest officers / employees will be given the right to use guns to fight against the criminal gangs. 'Secret Fund' will be developed to handle such united forest crimes.

• The basic programme of village area development and participatory forest management will be strengthened to manage and protect the natural and man-made forests.

• Necessary regulations will be made about Saw mills to avoid any effect of saw mills on the protection of forests.

• Survey and demarcation of all forest land will be completed to prevent the encroachment in forest area. Funds and facilities will be provided to update all forest maps & forest land records.

• Scientific management of forests will be done to increase the plant survival rate and ultimately achieve the increase in productivity of natural as well as man-made forests.

• Exotic tree species will be used for plantation only after their long term study about their suitability in local environment, any effect caused by them on native tree species etc.

• Mangrove cover under government jurisdiction will be identified through satellite images and their maps on suitable scales will be prepared by Forest Department.

• Mangrove forest having minimum manageable area will be taken up by forest department and their scientific management will be done with the help of local people.

• Special attention will be given to protect ecologically sensitive area and scattered forest areas.

• Technical and professional help will be given by forest department for development of private forests returned to the owner. Management of private forest undertaken by government will be done according to approved work plan.

• Compensation and financial support will be provided to the family of forest officer/ employee if he is handicapped or killed during completing his duty of forest protection.

Management of Catchment areas and Soil Conservation

• Catchment area development programme will be implemented effectively to prevent the soil erosion due to rainwater and to arrest as much rainwater as possible. Water bodies created through this will be used to increase the productivity of forests and to irrigate nearby agricultural lands which will help in improving the standard of living of rural people.

• Soil and water conservation works will be taken up to arrest the heavy runoff of water and recharge the ground water.

Wildlife and Biodiversity Conservation

• Special attention will be given to wildlife protection and management, biodiversity conservation in Protected Areas (PAs) & areas outside PAs. Management plan of these areas will include wildlife research, environmental awareness, wildlife training, habitat management and environment protection. Ex-situ conservation of flora and fauna will be done through establishing botanical gardens and zoological parks.

• Measures will be taken to decrease the wildlife-human conflict and thereby providing necessary environment for wildlife management. Area under PAs will be increased to make it five per cent of the total geographical area of the State. Redemarcation of boundaries of PAs will be done without any effect on biodiversity conservation therein.

• Anti-poaching squads will be formed in areas sensitive for wildlife poaching like Mumbai, Thane, Pune, Nagpur, Chandrapur, Amaravati, Kolhapur. Secret fund will be formed to obtain confidential information about illegal wildlife trade and prevent it.

• Rehabilitation of local people in PAs will be done by consultation with them and compensation for loss of crop, domestic animals and any human death due to wildlife attack, will be done as fast as possible.

• Revenue generated in any PA will be utilized for its management and development, also for benefiting local people near PAs.

Forest Research

• 'Forest Research Action Plan' will be prepared for State to increase the productivity of natural and man-made forests, to identify natural regeneration areas, to choose genetically important saplings, to establish seed production centre, to preserve important tree species etc.

• Similarly research activities will be undertaken for identifying native plant species for plantation in different areas, developing plan for restoration of wastelands, preparing bio-fertilizers and bio-manures and their use, medicinal plant conservation etc.

Human Resource Development & Training

• ‘Human Resource Development Plan’ will be implemented in Forest Department, Social Forestry Department and FDCM to attract educated and skilled people to these departments.

• Forest training institutes in the State will be technologically upgraded and strengthened. Field officers, forest guards will be trained here and refresher courses for forest officers will also be undertaken.

Tribal and Forests

• Participation of tribal in Forest Protection, restoration and development will be emphasized. Similarly they will be involved in NTFP management in forest area. Industries will be promoted for increase in NTFP production and their trade. Suitable mechanism will be set for fare pricing and sale of forest based products from tribal. Special measures will be taken with the help of Tribal Development Department for improving financial conditions of tribal through their active participation in NTFP collection, storage, processing, standardization and trade.

• According to the 73rd amendment done in the Constitution, forest based tribal and locals will not be deprived of their traditional rights and facilities.

Joint Forest Management

• Participatory Forest Management will be strictly implemented as successful implementation of any plan / programme for ‘protection of natural forest/man-made plantations, restoration activities on wastelands’ depends on basic concept of Participatory Forest Management.

• As per the 73rd amendment done in the Constitution, Participatory Forest Management will be given the status of law. Active participation of people and NGOs will be emphasized for ‘Forest Development’ to become a peoples’ programme. Villages will be made aware of the benefits of joint forest management through use of mass communication media.

Eco-tourism

Ecotourism will be encouraged for its benefit to forest based locals and people in general. It will provide employment to local people and create awareness about nature conservation. Accordingly promotion of ecotourism in PAs and forests outside PAs will be done as per the work enlisted in State Eco-tourism Policy. Forest based Industry: Forest based industries will be encouraged to establish relation with farmers and cooperative organizations doing plantations. They will take responsibility of the produce from farmers; will provide good quality saplings to them.

Management of Information System

• ‘Management Information System’ will be developed to prepare work plans of different schemes, to achieve high quality managerial coordination, to monitor various aspects and for easy availability of basic information needed for evaluation of completed tasks. E-governance and transparent work system will be achieved through this.

• Surveying of forest resources and collection and publication of matter, maps needed for forest management will be done on priority. Satellite imagery and GIS technology will be used for forest cover mapping after fixed time interval and monitoring of forest resources.

Reduction in Use of Timber and Other Forest Produce

• Alternatives for timber and firewood will be promoted to reduce the stress on existing forests. Use of Bamboo (obtained from plantation) in place of timber wood, use of biogas instead of firewood, use of solar energy will be encouraged.

• Stall feeding of fodder will be practiced to check the uncontrolled grazing in forests. Other departments working in similar areas will be coordinated, like Rural Development Department, Animal Husbandry Department, Non-conventional Energy Sources department.

• Use of modern electric cremation system will be emphasized to decrease the use of wood for burning dead bodies. Buildings and Infrastructure Development: Residential and official buildings will be developed on BOT (Build Operate Transfer) basis, on land under Forest Department (non-forest land or forest land if necessary).

Financial Support for Forestry Sector

• Forestry sector has important contribution in providing employment to rural people, increasing their agricultural produce and maintaining environmental balance. Therefore provision of enough funds for this sector is necessary.

• 'Green Fund' will be established to invest in wasteland development, nature restoration and catchment area development. Money obtained through Clean Development Mechanism (CDM) under Kyoto Protocol will be used for "Green Fund". Similarly tax obtained through compensatory afforestation and forest development will be utilized for green fund. Thought will be given to taxing 2 % Green Cess to municipalities for using water from forest areas and then the money will be used for green fund.

• Green fund will be utilized for completion of area /environment development and afforestation activities.

Unit-16

2.9 Tourism Industry and Environment: Issues and Challenges

Tourism has become the second largest industry in the world in terms of creating employment and generating income in many countries, it is a major source of foreign exchange earner. The economy of many countries mainly depends on International Tourism. Tourism contributes a very high percentage of G.D.P of those countries. France, Italy, Malaysia, UK, USA, and Singapore etc. these countries are most popular in tourist destination.

For example, Maldives mainly depends on International Tourism. There are no industries and nothing is grown there. The significance of tourism has been recognized in both developed and developing countries. There is wide spread optimism that tourism might become a powerful and beneficial agent of both economic and social change. Indeed, tourism has really stimulated both investment and employments. Today tourism is the largest service industry in India, with a contribution of 6.23% to the National GDP164 and providing 8.78% of the total employment.

India witness's more than 5 million annual foreign tourist arrivals and 562 million domestic tourism visits. The tourism industry in India generated about US$100 billion in 2008 and that is expected to increase to US$275.5 billion by 2018 at a 9.4% annual growth rate. The Ministry of Tourism is the nodal agency for the development and promotion of tourism in India and maintains the "Incredible India" campaign.

According to World Travel and Tourism Council, India will be a tourism hotspot from 2009-2018, having the highest 10-year growth potential. As per the Travel and Tourism Competitiveness Report 2009 by the World Economic Forum, India is ranked 11th in the Asia Pacific region and 62nd overall, moving up these places on the list of the world's attractive destination. It is ranked the 14th best tourist destinations for its natural resources and 24th for its cultural resources, with many World Heritage Sites, both natural and cultural, rich fauna, and strong creative industries in the country. India also bagged 37th rank for its air transport network. India travel and tourism industry ranked 5th in the long-term (10-year) growth and is expected to be the second largest employer in the world by 2019.

The tourism industry is one of the main sectors in the global economy, and it is often referred to as the world's largest single industry. In addition, it is one of the world's fastest growing industries and a major source of employment and foreign exchange even in many developed countries, with a focus on natural environments. Harnessing the opportunities and dealing with the challenges of the largest on-going migration of people is of utmost importance, and is particularly significant for the developing countries.

Tourism and its Impact on Environment

Like most of the human activities, tourism also has positive and negative impact on the society and environment as a whole. For the purpose of the research, it is important to assess and examine both the positive and negative impacts in a neutral way. The quality of the environment, both natural and man-made, is essential to tourism. However, tourism's relationship with the environment is complex178. It involves many activities that can have adverse environmental effects. Many of these impacts are linked with the construction of general infrastructure such as roads and airports, and of tourism facilities, including resorts, hotels, restaurants, shops, golf courses and marinas. The negative impacts of tourism development can gradually destroy the environmental resources on which it depends. On the other hand, tourism has the potential to create beneficial effects on the environment by contributing to environmental protection and conservation. It is a way to raise awareness of environmental values and it can serve as a tool to finance protection of natural areas and increase their economic importance.

Positive Impact of Tourism

a. Generating Income and Employment

Tourism in India has emerged as an instrument of income and employment generation, poverty alleviation and sustainable human development. It contributes 6.23% to the National GDP and 8.78% of the total employment in India. Almost 20 million people are now working in the India's tourism industry.

b. Source of Foreign Exchange Earnings

Tourism is an important source of foreign exchange earnings in India. This has favourable impact on the balance of payment of the country. The tourism industry in India generated about US$100 billion in 2008 and that is expected to increase to US $ 275.5 billion by 2018 at a 9.4% annual growth rate .

c. Preservation of National Heritage and Environment

Tourism helps preserve several places which are of historical importance by declaring them as heritage sites. For instance, the Taj Mahal, the Qutab Minar, Ajanta182 and Ellora temples, etc., would have been decayed and destroyed had it not been for the efforts taken by Tourism Department to preserve them. Likewise, tourism also helps in conserving the natural habitats of many endangered species.

d. Developing Infrastructure

Tourism tends to encourage the development of multiple-use infrastructure that benefits the host community, including various means of transports, health care facilities, and sports centres, in addition to the hotels and high-end restaurants that cater to foreign visitors. The development of infrastructure has in turn induced the development of other directly productive activities.

e. Promoting Peace and Stability

Honey and Gilpin (2009) suggests that the tourism industry can also help promote peace and stability in developing country like India by providing jobs, generating income, diversifying the economy, protecting the environment, and promoting cross-cultural awareness. However, key challenges like adoption of regulatory frameworks, mechanisms to reduce crime and corruption, etc., must be addressed if peace-enhancing benefits from this industry are to be realized.

f. Improved Environmental Management and Planning

Sound environmental management of tourism facilities and especially hotels can increase the benefits to natural environment. By planning early for tourism development, damaging and expensive mistakes can be prevented, avoiding the gradual deterioration of environmental assets significant to tourism. The development of tourism has moved the Indian Government towards this direction leading to improved environmental management.

g. Creating Environmental Awareness

Tourism has the potential to increase public appreciation of the environment and to spread awareness of environmental problems when it brings people into closer contact with nature and the environment. This confrontation heightens awareness of the value of nature among the community and lead to environmentally conscious behaviour and activities to preserve the environment.

Negative Impact on Environment

Tourism can cause the same forms of pollution as any other industry are emissions, noise, silid waste and littering, releasing of sewage, oil and chemicals, even architecture/visual pollution.

a. Depletion of Local Resources

Tourism can create pressure on local recourses like energy, food, and transport of these resources exacerbates the physical impacts associated with their exploitation. Because of the seasonal character of the industry, many destinations have ten times more inhabitants in the high season. A high demand is placed upon these recourses to meet the high expectorations tourists often have proper heating, hot water, etc.

b. Land degradation

Important land recourses include minerals, fossil fuels, fertile soil, forests, wetland and wildlife. Increases construction of tourism and recreational facilities has increased the pressure on these recourses and scenic landscapes. Direct impact on natural recourses, both renewable and non-renewable, in the provision of tourist facilities is caused by the use of land for accommodation and other infrastructure provision, and the use of building materials.

c. Water Pollution

Water especially, is one of the most critical natural resources. The tourism industry generally overuses water resources for hotels, swimming fools, golf courses and personal use of water by tourists. This can result in water shortages and degradation of water supplies, as well as generating a greater volume of wastewater186. In dryer regions like Rajasthan, the issue of water scarcity is particular concern. If a proper sewage disposal system has not been installed for hotels, resorts and other tourist facilities, there may be pollution of ground water from the sewage, or if a sewage outfall has been construed into a nearby river, lake or coastal seawater and the sewage has not been adequately treated, the effluent will pollute that water area.

This situation is common in beach resort areas where the hotels construct an outfall into the adjacent water area, which can also be used for swimming by tourists or for fishing by locals. Recreational and tourist transportation motor boats in surface water result in pollution in river, lakes and sea water due to spilling oil and gas and cleaning their bilge in water. This is usually common in enclosed harbour and places where natural water circulation is slow. Water resources are a prime attraction for tourism and recreational developments, and thus suffer impacts.

Water pollution is a result of waste water generated by tourist facilities and runoff, and occurs on inland lakes and streams as well as in the marine environment. Much of this is non-point pollution such as septic tank seepage, lawn fertilizer, road oil and runoff from disturbed soil. An extra nutrient in the water system causes eutrophication of lakes and streams, which in turn influences other aquatic life. Lakes choked with weeds and beaches with algae, a process accelerated by human influence, have become common in some areas. Inadequately treated effluent or raw sewage discharged into in water resources is a health hazard, as well. Water pollution is an increasingly serious problem in some areas such as the Mediterranean.

d. Air Pollution

Tourism is generally considered a “smokeless industry”. However, it can also result in air pollution by tourist vehicles in a particular area, especially at major attraction sites, that are accessible only by road. This is due to improperly maintained exhaust systems of the vehicles. In addition, pollution in the form of dust and dirt in the air may be generated from open, devegetated area if the tourism development is not properly planned, developed and landscaped or is in an interim State of construction. Transport by Air, road, and rail is continuously increasing in response to the rising to the rising number of tourists activities in India. Transport emissions and emissions from energy production and use are linked to acid rain, global warming and photochemical pollution.

Air pollution from tourists’ transportation has impacts on the global level, especially from carbon dioxide (CO2) emissions related to transportation energy use. In addition, it can contribute to severe local air pollution. Some of these impacts are quite specific to tourist activities where the sites are in remote areas like Ajanta and Ellora temples. For example, tour buses often leave their motors running for hours while the tourists go out for an excursion because they want to return to a comfortably air-conditioned bus.

e. Noise Pollution

Noise generated by a concentration of tourists road and certain types of tourist attractions such as amusement parks or car/motorcycle race tracks may reach uncomfortable and irritating levels for nearby residents and other tourists. Such loud noise can often result in ear damage and psychological stress. Noise pollution from airplanes, cars, and buses, as well as recreational vehicles is an ever-growing problem of modern life. In addition to causing annoyance, stress, and even hearing loss for humans, it causes distress to wildlife, especially in sensitive areas

f. Visual Pollution

It may result from several sources. These can be due to poorly- designed hotels and other facility buildings, which are not compatible with local architectural style or not well integrated into the natural environment. Other reasons can be poor maintenance of buildings and landscaping obstruction of scenic views by development use of large and ugly advertising signs. Littering of landscape also results in visual pollution.

g. Aesthetic Pollution

Often tourism fails to integrate its structures with the natural features and indigenous architectural of the destination. Large, dominating resorts of disparate design can look out of place in any natural environment and may clash with the indigenous structural design. A lack of land-use planning and building regulations in many destinations has facilitated sprawling developments along coastlines, valleys and scenic routes. The sprawl includes tourism facilities themselves and supporting infrastructure such as roads, employee housing, parking, service areas, and waste disposal.

h. Waste Disposal Problems

The most common problem in tourism areas is the littering of debris on the landscape. This is due to large number of people using the area of picnicking. Improper disposal of solid waste from hotel restaurants, and resorts generate both litter and environmental health problems from vermin, disease and pollution. It can also lead to the degradation of tourist sites. In areas with high concentrations of tourist's activities and appealing natural attractions, waste disposal is a serious problem and improper disposal can be a major despoiler of the natural environment-rivers, scenic areas, and roadsides. In mountain areas of the Himalayas and Darjeeling their garbage, oxygen cylinders and even camping equipment. Such practices degrade the environment particularly in remote areas because they have few garbage collection or disposal facilities.

i. Sewage

Construction of hotels, recreation and other facilities often leads to increased sewage pollution. Wastewater has polluted seas and lakes surrounding tourist attractions, damaging the flora and fauna. Sewage runoff causes serious damage to coral reefs because it stimulates the growth of algae, which cover the filter feeding corals, hindering their ability to survive. Changes in salinity and siltation can threaten the health of humans and animals. Examples of such pollution can be seen in the coastal States of Goa, Kerala, Maharashtra, Tamil Nadu, etc.

j. Destruction and Alteration of Ecosystem

An ecosystem is a geographic area including the entire living organism (people, plants, animals, and microorganisms) their physical surroundings such as soil as soil, water, and air) and the natural cycles that sustain them. Attractive landscape sites, such as sandy beaches in Goa, Maharashtra, Kerala, TamilNadu, lakes, riversides, and mountaintops and slops, are often transitional Zones, characterized by species-rich ecosystems. The threats to and pressures on these ecosystems are often severe because such places are very attractive to both tourists and developers. Examples may be cited from Krushedei Island near Rameswaram. What was once called paradise for marine biologists has been abandoned due to massive destruction of coral and other marine life. Another area of concern, which emerged at jaisalmer, is regarding the deterioration of the desert ecology due to increased tourist activities in the desert.

k. Undesirable Social and Cultural Change

Tourism sometimes led to the destruction of the social fabric of a community. The more tourists coming into a place, the more the perceived risk of that place losing its identity. A good example is Goa. From the late 60's to the early 80's when the Hippy culture was at its height, Goa was a haven for such hippies. Here they came in thousands and changed the whole culture of the State leading to a rise in the use of drugs, prostitution and human trafficking. This had a ripple effect on the country.

l. Creating a Sense of Antipathy

Tourism brought little benefit to the local community. In most all-inclusive package tours more than 80% of travellers' fees go to the airlines, hotels and other International companies, not to local businessmen and workers. Moreover, large hotel chain restaurants often import food to satisfy foreign visitors and rarely employ local staff for senior management positions, preventing local farmers and workers from reaping the benefit of their presence. This has often created a sense of antipathy towards the tourists and the government. Sociology is the science of society, social institutions, and social relationships. Visitors to a community or area create social relationships that typically differ greatly from the affiliations among the indigenous population.

Self-assessment Questions

1) Describe the components of an ecosystem.
2) Define ecological pyramids and describe with examples, pyramids of number and biomass.
3) What is primary productivity? Give brief description of factors that affect primary productivity.
4) Define decomposition and describe the processes and products of decomposition.
5) Give an account of energy flow in an ecosystem.
6) Write important features of a sedimentary cycle in an ecosystem.
7) Outline salient features of carbon cycling in an ecosystem.
8) Define the term ecology.
9) What are the levels of biological organizations at which ecological interactions occur?
10) Discuss the areas that ecology is concerned.
11) Have you observed, in your locality, any changes in pattern of disease occurrence that are related to changes in the environment?
12) What is environmental disease? Give examples in your setting.
13) What is Environment? Discuss the scope of Environment.
14) Describe the importance of environment studies.
15) What is an ecosystem?

16) What do you mean by natural environment?

17) Which are the major components of the environment?

18) Give four examples of human made environment

19) “The need for public awareness about environment is of vital importance.” Discuss.

20) Discuss the various types of environment.

21) Write an essay on the Silent Valley Movement.

22) How do you look at the Silent Valley Movement from the angle of the new social movements?

23) List at least FIVE reasons why a company might wish to undertake an environmental audit.

24) Write the negative impacts of tourism on the environment.

References

Bruntland et al 2012: Environment and Development Challenges: The Imperative to Act, Blue Planet Prize, The Asahi Glass Foundation.

Buckley M. (2011) John Stuart Mill and the Idea of a Stationary State Economy. In: Dierksmeier C., Amann W., von Kimakowitz E., Spitzeck H., Pirson M. (eds) Humanistic Ethics in the Age of Globality. Humanism in Business Series. Palgrave Macmillan, London.

Chakravarty, S. S.K. Ghosh, C.P. Suresh, A.N. Dubey & G. Shukla (2012). Deforestation: Causes, Effects and Control Strategies in Okia, C.A. (ed.), *Global Perspectives on Sustainable Forest Management* , Shanghai, China: InTech.

Chapman J.L. and Reiss, M.J. 1993: Ecology: Principles and Applications, Cambridge University Press, Cambridge.

Chopra, R. 2016: Environmental Degradation in India: Causes and Consequences, International Journal of Applied Environmental Sciences, 11(6), p. 1593-1601.

Contreras-Hermosilla, A. (2000). The Underlying Causes of Forest Decline, Centre for International Forestry Research, Occasional Paper No. 30.

Dash, M.C., 1993: Fundamentals of Ecology, Tata McGraw-Hill, New Delhi.

Handl, G. 2012: Declaration of the United Nations Conference on the Human Environment (Stockholm Declaration), 1972 And The Rio Declaration On Environment And Development, 1992, United Nations Audiovisual Library of International Law.

Ingole, S. P. 2012: Environmental Auditing: Its Benefits and Counterance, *International Journal of Science Innovations and Discoveries*, 2 (5)(2), p. 152-6.

Järvilehto, T. 2009: The Theory of the Organism-Environment System as a Basis of Experimental Work in Psychology, Ecological Psychology, 21 (2), p. 112-120.

Kinda, S. R. 2013: Essays on environmental degradation and economic development. Economics and Finance. Université d'Auvergne - Clermont-Ferrand I, English. ffNNT : 2013CLF10411ff.

Kőszegi, M. Bottlik, Z. Telbisz, T. and Mari, L. 2015: Human-environment relationships in modern and postmodern geography, Hungarian Geographical Bulletin, 64(2), p. 87-99.

Legesse, W., Mulugeta, T. & Ambelu A. 2002: Introduction to Ecology. EPHTI.

Odum, E.P. 1997: Ecology: A Bridge between Science and Society, Sinaur Associates Inc. Publishers, Sunderland.

Sharma, P.D. 1996: Ecology and Environment, 71h edition, Rastogi Publications, Mirat.

Watson, E.E., Simmons, I., Fernández-Armesto, F. and Sluyter, A. 2010: Environmentalist Thinking and/in Geography, Progress in Human Geography, 34(1) p. 98–116.

Webliography

1) http://ncert.nic.in/ncerts/l/lebo114.pdf

2) https://www.learner.org/courses/envsci/unit/pdfs/unit4.pdf

3) https://sc-s.si/joomla/images/Ecosystems.pdf

4) http://download.nos.org/333courseE/5.pdf

5) https://ocw.mit.edu/courses/civil-and-environmental-engineering/1-018j-ecology-i-the-earth-system-fall-2009/lecture-notes/MIT1_018JF09_Lec01.pdf

6) http://www.biodiversitybc.org/assets/pressReleases/BBCPrinciplesWEB.pdf

7) http://www.bio.miami.edu/ecosummer/lectures/lec03_climate.pdf

8) http://www.yourarticlelibrary.com/environment/environmental-studies-meaning-objectives-scope-and-importance-of-environmental-studies/12295

9) https://www.quora.com/What-are-the-concept-and-scope-of-environment-geography

10) https://www.tutorialspoint.com/environmental_studies/environmental_studies_tutorial.pdf

11) https://courses.seas.harvard.edu/climate/eli/Courses/EPS281r/Sources/Gaia/Gaia-hypothesis-wikipedia.pdf

12) http://www.religionandnature.com/ern/sample/Taylor+Zimmerman--DeepEcology.pdf

13) http://www.uky.edu/OtherOrgs/AppalFor/Readings/240%20-%20Reading%20-%20Deep%20Ecology.pdf

14) https://www.uwosh.edu/facstaff/barnhill/ES-243/pp%20outline%20Deep%20Ecology.pdf

15) https://en.wikipedia.org/wiki/Spaceship_Earth

16) https://www.researchgate.net/publication/279201881_Environmental_Degradation_Causes_Impacts_and_Mitigation

17) https://shodhganga.inflibnet.ac.in/bitstream/10603/165832/14/14_chapter_05.pdf

18) http://egyankosh.ac.in/bitstream/123456789/19894/1/Unit-11.pdf

19) http://www.unesco.org/education/pdf/RIO_E.PDF

20) https://en.wikipedia.org/wiki/Montreal_Protocol

21) https://ozone.unep.org/sites/default/files/Consolidated-Montreal-Protocol-November-2016.pdf

22) https://treaties.un.org/doc/publication/unts/volume%201522/volume-1522-i-26369-english.pdf

23) https://associationdatabase.com/aws/OAIMA/asset_manager/get_file/12904?ver=17581

24) https://www3.opic.gov/environment/eia/kipeto/EMP/KEL%20Transmission%20Line%20Environment%20Management%20Plan.pdf

25) http://environmentclearance.nic.in/writereaddata/EIA/30052017MITJ5FG8Annexure-documentofEIA_EMP.pdf

26) http://environmentclearance.nic.in/writereaddata/EIA/30052017MITJ5FG8Annexure-documentofEIA_EMP.pdf

27) http://mahenvis.nic.in/Pdf/Forests_.pdf

28) https://shodhganga.inflibnet.ac.in/bitstream/10603/97098/11/11_chapter%204.pdf

Disclaimer

This self-learning material is based on different books, journals and web-sources.

Self-Learning Material (SLM) for the Course of

M.A. / M.Sc. in Geography

Directorate of Open and Distance Learning (DODL)

University of Kalyani

Paper: GEO311T (FG)

(Total Credit – 4; Total Marks – 100: Internal Evaluation – 20 + Semester-end Examination - 80)

University of Kalyani

Kalyani

Nadia – 741235

West Bengal

Course Material Compiled by:

Dr. Sayan Choudhary

Assistant Professor (Contractual)

Directorate of Open and Distance Learning

University of Kalyani, West Bengal

December 2019

Directorate of Open and Distance Learning, University of Kalyani.

Published by the Directorate of Open and Distance Learning,

University of Kalyani, Kalyani-741235, West Bengal.

Director's Message

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distance Learning (DODL), University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC Regulations, 2017 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of the SLMs, making them useful to the learners, besides avoiding errors as far as practicable. Further suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from Professor (Dr.) Sankar Kumar Ghosh, Hon'ble Vice- Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticism to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every member of PGBOS (DODL), University of Kalyani. Heartfelt thanks is also due to the Course Writers-faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and co-ordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the curriculum requirements of the Post Graduate Programme through Distance Mode.

All the Self Learning Materials are self writing and collected from e-book, journals and websites.

Director
Directorate of Open and Distance Learning
University of Kalyani

SYLLABUS

Paper –XI: Special Paper Theory (Fluvial Geomorphology*)*

Group–A: Basics of Fluvial Geomorphology [Credit: 2]

(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit. 01	Scope and significance of fluvial geomorphology;
Unit. 02	Fluvial system: concepts, components, input output with stores of material and energy;
Unit. 03	Variables of fluvial system: internal and external, adjustable and controlling; Scales in fluvial geomorphology: times and space scales;
Unit. 04	Channel initiation: theory of overland flow, theory of sub-surface flow;
Unit. 05	Flow regime: hydrological pathways, measurement of stream flow and annual flow regimes; Classification of natural streams by D.L. Rosgen
Unit. 06	Drainage evolution: parallel development and coalescence, headward extension and branching, lateral expansion;
Unit. 07	Drainage network: composition, laws of stream number and stream length, number of Topologically Distinct Channel Networks, classification of link types;
Unit. 08	Analysis of drainage basin: Basin shape, form factor (Horton, 1932), elipticity index (Stoddart, 1965), circularity index (Miller, 1953), influence of basin shape on hydrological regime; Basin size: area ratio and law of basin area, law of allometric growth and basin area; Relation of basin area with stream number and length;

Group – B: Mechanism of Fluvial Processes [Credit: 2]

(Marks - 50: Internal Evaluation – 10, Semester-end Examination - 40)

Unit. 09	Mechanics of flow: uniform and non-uniform, steady and unsteady, laminar and turbulent, tranquil and rapid;
Unit. 10	Velocity and resistance: factors affecting velocity, flow resistance and viscosity;
Unit. 11	Stream power and energy;
Unit. 12	Mechanics of fluvial erosion: threshold of erosion, entrainment and bed erosion, headward erosion and channel lengthening, bank erosion by fluvial processes;
Unit. 13	River transportation: process of particle entrainment, initiation of sediment transport, selective theories of sediment transport, theories of equal mobility transport, transport of bed load and suspended load;
Unit. 14	Sediment load: nature and types of sediment load, sources of sediment load;
Unit. 15	Mechanics of sediment deposit: Channel competence, capacity and efficiency;
Unit. 16	Sediment deposition - processes, nature and characteristics of fluvial deposits; Flood plain and deltaic plain deposits

PAPER –XI: SPECIAL PAPER THEORY (FLUVIAL GEOMORPHOLOGY*)*
GROUP–A: BASICS OF FLUVIAL GEOMORPHOLOGY [CREDIT: 2]
(MARKS - 50: INTERNAL EVALUATION – 10, SEMESTER-END EXAMINATION - 40)

1.1 INTRODUCTION

The following section discuss scope and significance of fluvial geomorphology, fluvial system, variables of fluvial system, channel initiation, flow regime, drainage evolution and drainage network. Analysis of various aspects of drainage morphometry will further enhance the students to ability to grasp different topics of fluvial morphometry.

1.2 LEARNING OBJECTIVES

The following topics will be discussed in the present section -

- Scope and significance of fluvial geomorphology;
- Fluvial system: concepts, components, input output with stores of material and energy;
- Variables of fluvial system: internal and external, adjustable and controlling; Scales in fluvial geomorphology: times and space scales;
- Channel initiation: theory of overland flow, theory of sub-surface flow;
- Flow regime: hydrological pathways, measurement of stream flow and annual flow regimes; Classification of natural streams by D.L. Rosgen
- Drainage evolution: parallel development and coalescence, headward extension and branching, lateral expansion;
- Drainage network: composition, laws of stream number and stream length, number of Topologically Distinct Channel Networks, classification of link types;
- Analysis of drainage basin: Basin shape, form factor (Horton, 1932), elipticity index (Stoddart, 1965), circularity index (Miller, 1953), influence of basin shape on hydrological regime; Basin size: area ratio and law of basin area, law of allometric growth and basin area; Relation of basin area with stream number and length

1.3 ASSESSMENT OF PRIOR KNOWLEDGE

Fundamental concepts of geomorphology will be discussed.

1.4 LEARNING ACTIVITIES

Identification and demarcation of drainage basin will be done.

1.5 FEEDBACK OF LEARNING ACTIVITIES

Discussion on various aspects of fluvial morphometry will be helpful.

1.6 EXAMPLES AND ILUUSTRATION

Unit – 1: SCOPE AND SIGNIFICANCE OF FLUVIAL GEOMORPHOLOGY

When a man makes a pilgrimage to the fields and woods of his boyhood, he does not expect to find the hills and mountains dissolved, or the valleys moved. If other men have not tom up the land to build factories and towns, he expects his children to see the hills and swales as his forefathers saw them. And he *is* almost right. Probably neither he nor the children will ever notice that in fifty years the surface of the ground has been lowered perhaps a fraction of an inch. Why should they? But they might not be surprised to find that the old mill pond behind the dam is now more mud than water. Under the action of the force of gravity the land surface is sculptured by water, wind, and ice. This sculpturing produces the landforms with which geomorphology is concerned. Some of these forms owe their origins purely to denudational processes; other forms may be depositional; still others owe their existence to combinations of both processes.

A picture of the dynamics of the earth's surface is by no means complete, however, if only gradation or leveling is considered. Clearly, if there were no counteracting forces we should expect that the land surface, given sufficient time, would be continuously reduced. Eventually, little or no relief would remain. Geologic history demonstrates, however, that the degradational forces acting on the earth's surface are opposed by constructional forces. These internal, or endogenous, forces cause the land to rise, and as they do so it is subjected to attack by the external, or exogenous, agents.

Geomorphology is primarily concerned with the exogenous processes as they mold the surface of the earth, but the internal forces cannot be disregarded when one considers fundamental concepts of the origin and development of landforms. Ideally, the basic principles underlying the development of landforms can be considered *in* simple terms. A given land area is composed of a particular set of rocks, which have particular chemical and mineralogic compositions and specific physical properties. Because these rocks were formed at different temperatures and pressures within the earth, when they are exposed at the surface they are no longer in equilibrium with their environment and thus begin to decompose. Where a gradient is created by gravity, the moving water, earth, air, and ice help in the attack upon the rock and remove the products of weathering. In the process, landforms of various aspects are created. In a given environment the physical and chemical constitution of the rocks determines the way in which they will break down and, in turn, the size and quantity of debris made available to the denudational agencies.

Each denudational agent, depending upon its density, gradient, and mass at a particular place, is capable of applying a given stress on the materials available. A certain amount of work may be performed by the application of this stress, and the results of this work are the landforms that we see developed in various parts of the world. In a given climatic and vegetational environment the shape or form of the landscape will vary, depending upon the character of the rock and the type and available stress of the erosional agents. But as the land surface is reduced-so long as the products of weathering and the applied stress remain constant-the form of the land should remain the same.

If one were able to evaluate properly the properties of the rocks and the present and past capabilities of the denudational agencies, he should have no trouble in developing a rational, even mathematical, equation capable of describing the development history and equilibrium form of any landscape. William Morris Davis said essentially the same thing in 1902 when he observed that any landform is a function of the structure of the rocks (including their composition and structural attitude), the processes acting upon them, and the time over which these processes have been active. Only as we study the interrelations of these three factors are we able to discern which combinations produce which particular landforms and how they do so.

Some landforms, such as volcanoes, which may have been unaffected by denudational processes, may be considered purely constructional forms. As soon, however, as they are modified by external agencies, their form begins to represent the resultant of an interaction between the constructional forces, the rock substrate, and the applied stress. The application of such an ideal concept to any actual landform at the present time is fraught with problems. The natural world is highly variable and the mechanics of uplift, weathering, and erosion are for the most part poorly understood. As will be seen, climate itself is a complex factor, and in most regions of the world inorganic processes are inseparable from the complex organic processes carried on by plants and animals. Although it is frequently convenient and helpful to construct a simplified synthetic picture of the natural environment, we should not lose sight of the fact that a given landscape must be the result of a complex set of factors which encompass the behavior of materials and processes over varying periods of time.

It is important to note that whether one refers to the effect on landforms of different rock types, or to the effect of different rates of uplift, such differences or changes must manifest themselves in the environment of the landform in simple physical terms. A normal fault whose strike is perpendicular to the direction of flow of a river, with downthrown block in the downstream direction, constitutes to the river a merely local increase in gradient. A similar increase in gradient might be effected by local changes in lithology, an abrupt shortening in channel length, or by an abrupt change in discharge downstream. The same physical principles determine the river's subsequent response in each case. The permanence or impermanence of the change, as well as its possible propagation either upstream or down, will depend upon the type and amount of material available and the distribution and quantity of flow. Any true principle enunciated to explain one of the cases must be applicable to the others as well.

Thus, although the application of the principle to anyone example may be fraught with difficulty, an understanding of the principle at least reduces the burden of innumerable "unique" cases.

Geomorphologists have always sought such unifying concepts, and for a proper view of the field as a whole one must turn initially to the classical concepts of landform evolution.

The influence of William Morris Davis on geomorphology was without doubt greater and longer-lasting than that of any other individual. His major contribution was a genetic system of landform description. Beginning in 1899, Davis developed the concept that during erosion of a highland the landscape evolves systematically through distinctive stages, to which he gave the names, youth, maturity, and old age. This entire sequence of stages he called an erosion cycle (or geomorphic cycle), and the end product was supposed to be a surface of low relief, or peneplain.

He elaborated the effects of interruptions in the cycle and argued that the principal factors controlling the character of landforms are geologic structure, geomorphic processes, and the stage of development. Davis' genetic concept of landform development was a brilliant synthesis, which grew directly out of the work by Powell, Gilbert, and Dutton and also from the controversial ideas on organic evolution which were prevalent at the time.

The concept of the erosion cycle was never accepted in Europe to the same degree as in North America. The most serious challenge came during the 1920's from Walther Penck, who attempted to show a direct causal relation between tectonics and the properties of landforms. Many of his conclusions about the trends and ultimate results of tectonics and erosion processes differed only slightly from those of Davis. Penck, however, emphasized slope development, and his theory of slope development is a major contribution that is still being tested and debated.

The principal alternative to the Davisian conception differs mainly in the view of the effect of time, the third of the three fundamental elements, on landforms. Restating and extending the work of Gilbert, Hack (1960) emphasizes the concept of a dynamic equilibrium in the landscape which is quickly established and which responds to changes that occur during the passage of time. This view postulates that there is at all times an approximate balance between work done and imposed load and that as the landscape is lowered by erosion and solution, or is uplifted, or as processes alter with changing climate, adjustments occur that maintain this approximate balance.

More will be said about these different views in subsequent chapters, as various aspects of the landscape are considered in greater detail. Paralleling developments in other phases of geology, the past decade has witnessed a remarkable increase in the application of analytical and experimental techniques to geomorphic problems. These investigations have taken two principal directions: (1) efforts to describe landforms more precisely through the use of statistics and other analytical techniques, (2) application of physical and chemical principles to field and laboratory studies of geomorphic processes. Although a few geologists-G. K. Gilbert, and later W. W. Rubey-helped to pave the way for this current trend, developments in other fields of science, especially in engineering and physics, were more directly responsible for it. One outstanding example is the field and experimental work on sand transport by R. A. Bagnold during the 1930's. Another is the contribution of fundamental ideas on the development of stream networks by R. E. Horton. Recently many developments in hydraulics and in the application of soil mechanics have attracted the attention of geomorphologists. At present there is greatly increased interest in the use of more precise tools for studying landforms. The pace of research seems to be quickening and there is reason to hope that a new era of discovery is under way.

Geomorphology in North America has gone through a phase during which extensive description of the landscape in terms of the erosion cycle has been carried out. It was apparently believed that the processes were known or could be inferred, and that form could be assessed by eye. Similarly, one current earth-history view of geomorphology assumes that enough is now known to interpret landforms and deposits in terms of processes that operated in times past. In the most qualitative way this is probably true. However, we believe that the genetic system breaks down when it is subjected to close scrutiny involving quantitative data. At present deductions are subject to considerable doubt, for the detailed properties of landform have not been studied carefully enough and the fundamental aspects of most geomorphic processes are still poorly understood.

So long as this is true, the interpretation of geomorphic history rests on an exceedingly unstable base. Accordingly, we plan to concentrate on geomorphic processes. The emphasis is primarily upon river and slope processes; river processes will receive greatest attention, since the greatest volume of information available is on rivers. Our objective is to synthesize the material on these subjects in an attempt to assess the current status of knowledge and at the same time to draw attention to its shortcomings.

Process implies mechanics-that is, the explanation of the inner workings of a process through the application of physical and chemical principles. We realize that some readers may be more interested in descriptions of landforms than *in* the detailed analysis of the processes that formed them. So far as possible, we attempt to relate the processes discussed to specific types of landforms. Unfortunately, the gap between our understanding of specific processes in microcosm and the explanation of major large-scale landforms is still wide. It is interesting to note that geomorphologists seem to have a better understanding of depositional than of erosional forms. This may be because the formation of depositional features such as sand dunes, deltas, and flood plains is more easily seen in the field, or because many erosional features retain less clear evidence of their mode of formation.

Detailed understanding of geomorphic processes is not a substitute for the application of basic geologic and stratigraphic principles. Rather, such understanding should help to narrow the range of possible hypotheses applicable to the explanation of different geomorphic forms and surficial earth processes and deposits.

Unit – 2: FLUVIAL SYSTEM: CONCEPTS, COMPONENTS, INPUT OUTPUT WITH STORES OF MATERIAL AND ENERGY

A system is a collection of related objects and the processes that link those objects together. Within fluvial systems, objects such as hillslopes, the channel network and floodplains are linked together by the processes that move water and sediment between them. In common with other systems, the fluvial system is hierarchical, in that there are integrated sub-systems operating within it.

INPUTS, OUTPUTS AND STORES

The basic unit of the fluvial system is the drainage basin. Fluvial systems are open systems, which means that energy and materials are exchanged with the surrounding environment. In closed systems, only energy is exchanged with the surrounding environment.

Inputs

The main inputs to the system are water and sediment derived from the breakdown of the underlying rocks. Additional inputs include biological material and solutes derived from atmospheric inputs, rock weathering and the breakdown of organic material. Most of the energy required to drive the system is provided by the atmospheric processes that lift and condense the water that falls as precipitation over the drainage basin. The pull of gravity then moves this water downslope, creating a flow of energy through the system. This energy is expended in moving water and sediment to river channels and through the channel network.

Outputs

Water and sediment move through the system to the drainage basin outlet, where material is discharged to the ocean. Not all rivers reach the ocean; some flow into inland lakes and seas, while others, such as the Okavango River in Botswana, dry up before reaching the ocean. This reflects another important output from fluvial systems: the loss of water by evaporation to the atmosphere. Most of the available energy is used in overcoming the considerable frictional forces involved in moving water and sediment from hillslopes into channels and through the channel network. Much of this energy is 'lost' to the atmosphere in the form of heat.

Stores

A certain amount of material is stored along the way. For example, water is stored for varying lengths of time in lakes and reservoirs, and below the ground in the soil and aquifers. Sediment is stored when it is deposited in channels, lake basins, deltas, alluvial fans and on floodplains. This material may be released from storage at a later stage, perhaps when a channel migrates across its floodplain, eroding into formerly deposited sediments which are then carried downstream. Ferguson (1981) describes the channel as 'a jerky conveyor belt', since sediment is transferred intermittently seawards.

UNIT – 3: VARIABLES OF FLUVIAL SYSTEM: INTERNAL AND EXTERNAL, ADJUSTABLE AND CONTROLLING; SCALES IN FLUVIAL GEOMORPHOLOGY: TIMES AND SPACE SCALES

TYPES OF SYSTEM

Three types of system can be identified in fluvial geomorphology. These are morphological systems, cascading systems and process–response systems.

Morphological (form) systems

Landforms such as channels, hillslopes and floodplains form a morphological system, also referred to as a **form system**. The form of each component of a morphological system is related to the form of the other components in the system. For example, if the streams in the headwaters of a drainage basin are closely spaced, the hillslopes dividing them are steeper than they would be if the streams were further apart from each other. Relationships such as this can be quantified statistically.

Cascading (process) systems

The components of the morphological system are linked by a cascading system, which refers to the flow of water and sediment through the morphological system. Cascading systems are also called **process systems** or **flow systems**. These flows follow interconnected pathways from hillslopes to channels and through the channel network.

Process–response systems

The two systems interact as a process–response system. This describes the adjustments between the processes of the cascading system and the forms of the morphological system. There is a two-way feedback between process and form. In other words, processes shape forms and forms influence the way in which processes operate (rates and intensity). This can be seen where a steep section of channel causes high flow velocities and increased rates of erosion. Over time erosion is focused at this steep section and the channel slope is reduced. Velocity decreases as a result, reducing rates of erosion. In order to examine the components of the fluvial system in more detail, it can be divided into sub-systems, each operating as a system within the integrated whole. One way of doing this is to consider the system in terms of three zones, each of which is a process–response system with its own inputs and outputs (Figure 2.1). Within each zone certain processes dominate. The sediment **production zone** in the headwater regions is where most of the sediment originates, being supplied to the channel network from the bordering hillslopes by processes of erosion and the mass movement of weathered rock material. This sediment is then moved through the channel network in the sediment **transfer zone**, where the links between the channel and bordering hillslopes, and hence sediment production, are not so strong. As the river approaches the ocean, its gradient declines and the energy available for sediment transport is greatly reduced in the sediment **deposition zone**. It is primarily the finest sediment that reaches the ocean, as coarser sediment tends to be deposited further upstream. In fact, only a certain proportion of all the sediment that is produced within a drainage basin actually reaches the basin outlet.

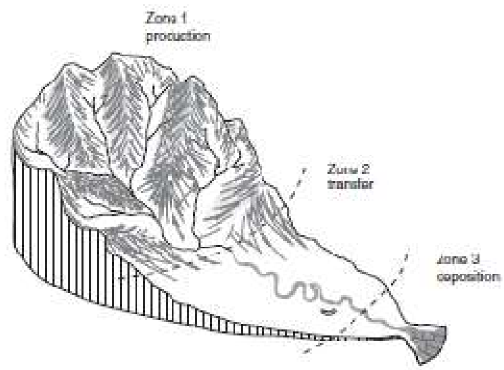


Figure 2.1 The fluvial system can be subdivided into three zones on the basis of the dominant processes operating within each zone. These are the sediment production zone, sediment transfer zone and sediment deposition zone. Adapted from Schumm (1977)

FLUVIAL SYSTEMS VARIABLES

Variables are quantities whose values change through time. They include such things as drainage density, hillslope angle, soil type, flow discharge, sediment yield, channel pattern and channel depth.

Internal and external variables

An important distinction exists between internal and external variables. All the examples given above are **internal variables**, which operate within the fluvial system. Internal variables are influenced by other internal variables, and also by variables that originate from outside the system. These **external variables**, such as climate, control or regulate the way in which the system operates. Unlike the internal variables, external variables operate independently, in that they are not influenced by what is going on inside the fluvial system. At the basin scale the external variables are climate, base level, tectonics and human activity. If you are considering a sub-system, such as a reach of channel in the transfer zone, the external variables would include the supply of flow and sediment to the channel. This is because these variables originate from outside the channel sub-system, even though they are internal variables at the basin scale. To avoid confusion, the 'ultimate' external variables – climate, base level, tectonics and human activity – will be referred to as **external basin controls**.

The external basin controls

The variables defined in this section act as regulators of the whole system. Any change in one of these variables will lead to a complex sequence of changes and adjustments within the fluvial system.

● ***Climate*** describes the fluctuations in average weather. Although the weather is always changing, longer-term characteristics such as seasonal and inter-annual variations can be defined. Other characteristics include how often storms of a given size can be expected to occur and the frequency and duration of droughts. Where no long-term changes are occurring in the climate, the combination of such attributes defines an envelope of 'normal' behaviour. **Climate change** occurs when this envelope shifts and a new range of climatic conditions arises.

● ***Tectonics*** refers to the internal forces that deform the Earth's crust. These forces can lead to large scale uplift, localised subsidence, warping, tilting, fracturing and faulting. Where uplift has occurred, inputs of water have to be lifted to a greater elevation, increasing energy availability; some of the highest rates of sediment production in the world are associated with areas of tectonic uplift. Valley gradients are altered by faulting and localised uplift, which may in turn affect channel pattern. Lateral (sideways) tilting can cause channel migration and affect patterns of valley sedimentation.

● ***Base level*** is the level below which a channel cannot erode. In most cases this is sea level. If there is a fall in sea level relative to the land surface, more energy is available to drive flow and sediment movement. Conversely a relative rise in base level means that less energy is available, resulting in net deposition in the lower reaches of the channel. Over time these effects may be propagated upstream through a complex sequence of internal adjustments and feedbacks.

● ***Human activity*** has had an increasing influence on fluvial systems over the last 5,000 years, especially during recent times. Activities within the drainage basin such as deforestation, agriculture and mining operations all affect the flow of water and production of sediment. These are referred to as indirect or diffuse activities. River channels are also modified directly when channel engineering is carried out. Advances in technology over the last century have meant that dam construction, channel enlargement for navigation and flood control, channel realignment, the building of flood embankments and other engineering works can now be carried out at an unprecedented scale. Today there are very few rivers that have not been affected in some way by the direct and indirect effects of human activity. It can be argued that, under some circumstances, human activity can be considered to be *both* an internal and an external variable. Many of the direct modifications described above are in response to some local human perception of the system. For example, channels are dredged because they are not deep enough for navigation, or flood defence works carried out because floods occur too frequently. Urban (2002) suggests that direct human intervention can often be classified as an internal variable, although it is more appropriate to consider *indirect* human activities as external.

The influence of the external basin controls on the fluvial system is represented schematically in Figure 2.2. Some internal variables have a greater degree of independence in that they are only affected in a limited way by the fluvial system. These variables are geology, soils and vegetation and topography (which includes relief, altitude and drainage basin size). All are internal variables because they are controlled to some extent by the external basin controls, however their main influence on the operation of the fluvial system is a controlling one.

Adjustable (dependent) and controlling (independent) variables

From the discussion above it can be seen that some variables control the adjustment of other variables. For example channel pattern is, among other things, affected by the supply of sediment to the channel.

In this case, channel pattern is the **adjustable** or **dependent variable** while sediment supply is the **controlling** or **independent variable**. Things can get a little confusing because controlling variables may in turn be adjusted by other variables. Extending the previous example, sediment supply is itself controlled by hillslope vegetation cover. In this case, sediment supply is the adjustable variable and vegetation cover the controlling variable. All internal variables are adjustable because their operation is ultimately regulated by the external basin controls. They are also influenced to a greater or lesser extent by other internal variables. Because the relationships between variables are so complicated, it can be very difficult to isolate the effect of one variable on another. The hierarchical nature of the fluvial system means that variables operating at larger scales tend to affect the operation of variables at smaller scales. For example, climate affects vegetation cover and hillslope erosion, which in turn determine sediment supply, which influences channel pattern, which affects the small-scale flow dynamics in the channel, which governs the movement of individual grains. This is not a one-way process, however.

Over long periods of time, the cumulative effect of small-scale processes, such as the erosion and deposition of individual grains, can lead to larger scale changes. These include changes in channel pattern and, over time periods of tens to hundreds of thousands of years, can adjust the slope of the whole river valley. Time itself is an important controlling variable. Every drainage basin has a historical legacy resulting from past changes that have taken place in the basin. This includes the cumulative effect of processes such as erosion, transport and deposition over long periods of time. It also includes the far-reaching effects of changes in the external basin controls, such as the variations in climatic conditions since the Last Glacial Maximum 18,000 years ago, which have greatly affected fluvial systems worldwide. In the temperate zone, many rivers underwent a transition from a braided to a meandering form as climate conditions ameliorated, vegetation became established and sediment loads decreased. However, vast quantities of sediment still remain in formerly glaciated drainage basins, where many fluvial systems are still adjusting to this glacial legacy.

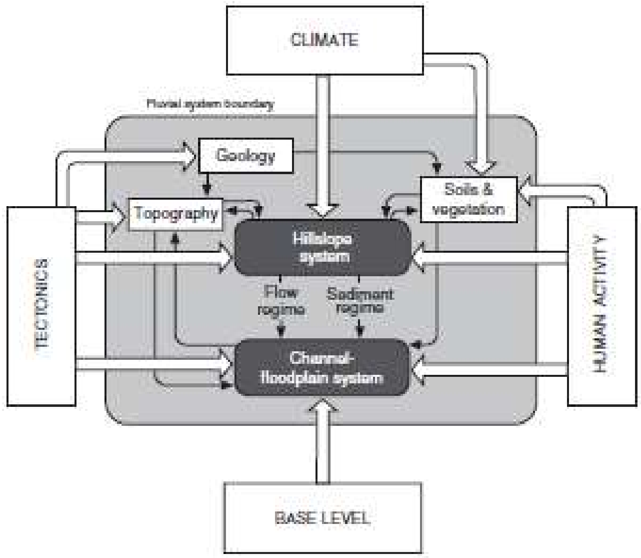


Figure 2.2 Simplified representation of the fluvial system.

Feedbacks

A **feedback** occurs when a change in one variable leads to a change in one or more other variables, which acts to either counteract or reinforce the effects of the original change. Two types of feedback are observed: negative feedback and positive feedback. Both are initiated by a change in one of the system variables, which in turn leads to a sequence of adjustments that eventually counteract the effect of the original change (**negative feedback**) or enhance it (**positive feedback**). When there is a change in one of the external controls, negative feedbacks allow the system to recover, damping out the effect of the change. An everyday example of a negative feedback loop is a central heating system controlled by a thermostat, which switches the source of heat on and off as the room cools and warms. An equilibrium is maintained as the temperature fluctuates around an average value. A commonly cited example of negative feedback within the fluvial system occurs when a section of channel is suddenly steepened by tectonic faulting. This leads to a local increase in the flow velocity and rate of bed erosion. Over time this acts to reduce the channel slope, counteracting the effects of the original change. It should be noted that the actual sequence of events is usually rather more complex. This is because change in one part of the system can lead to complex changes, both locally and throughout the rest of the system. The nature of complex response will be discussed later in this chapter. Positive feedbacks have a very different effect. Soil erosion is a natural process and an equilibrium exists if rates of soil removal over a given period are balanced by rates of soil formation over that period. However, an external change, such as the deforestation of steep slopes, can lead to a dramatic increase in soil erosion. The upper soil layers contain the most organic matter, which is important in binding the soil together. It also increases soil permeability, allowing rainfall to soak into the soil rather than running over and eroding the soil surface. If the topsoil is removed, the lower permeability of the underlying soil layers means that more water runs over the surface, increasing erosion and removing still more soil layers. In this way several centimetres of soil can be removed by a single rain storm (Woodward and Foster, 1997). This greatly exceeds the rate of soil formation. Referred to colloquially as ‘vicious circles’ or ‘the snowball effect’, positive feedbacks involve a move away from an equilibrium state. They usually involve the crossing of a **threshold** (see below) as the system moves towards a new equilibrium. A small-scale example of positive feedback is the build up of sediment during the formation of a channel bar. Bar formation is initiated when bedload sediment is deposited at a particular location on the channel bed (the various mechanisms by which this occurs will be discussed in later chapters). This affects local flow dynamics, causing the flow to diverge over and around the initial deposit. As the flow diverges, it becomes less concentrated and therefore less able to transport the coarser sediment. Localised deposition occurs, further disrupting the flow and promoting further deposition and bar growth. Several feedbacks, both positive and negative, exist between channel form, water flow and sediment transport. The form of a channel has an important influence on the way that water and sediment move through it. For example, flow is concentrated where the channel narrows, increasing erosion potential. As you saw above, deposition may occur where the flow diverges around obstacles such as bars. The character of the channel bed is also significant, since the size and arrangement of sediment determines bed roughness and resistance to flow. Where resistance is high, the average velocity of flow in the channel is reduced. This influences hydraulic conditions near the bed of the channel, which are significant for processes of erosion and deposition. Considerable differences are seen across the channel bed, giving rise to spatial variations in erosion and deposition. These processes themselves modify the form of the channel, feeding back to influence flow.

Thresholds

Thresholds are another important concept in systems theory and you will come across many examples in fluvial geomorphology. For example, a threshold is crossed when a sand grain on the bed of the channel is entrained (set in motion). Movement is resisted by the submerged weight of the grain and friction between it and the neighbouring grains. If the driving force exerted on the grain by the flow is less than these resisting forces, no movement will occur. It is only when the driving force of the flow exceeds the submerged weight of the grain that entrainment will take place. In this example, channel flow is an external variable. When a threshold is crossed there is a sudden change in the system, for example when loose material on a slope becomes unstable and starts to move down the slope as a landslide. The gradual processes by which rock is broken down and loose material builds up on a hillslope take place over time scales of tens to hundreds of years. Why, then, does a landslide occur at a particular point in time? Such a transition can come about when a change in one of the controlling variables leads to instability within the system – as a direct result of an earthquake (tectonics) for example. Thresholds that are crossed as a result of external change are called **external thresholds**. Instabilities may also develop over time without any external change having occurred. For this reason it is possible for a major landslide to be triggered by a relatively minor rainfall event that falls well within the expected climatic norm, because instability has gradually developed over time and the system is ready ‘primed’. This is an example of an **internal threshold**. Another commonly cited example is the threshold that is crossed when a meander loop is cut off to form an oxbow lake (several of these are shown in Colour Plate 11). Again this is something that can take place without there having been any change in the external variables. Whether or not either kind of threshold is crossed depends on how ‘sensitive’ the system is, in other words how close to a threshold it is. To illustrate this point, consider a pan filled with water that is heated by 10°C at normal atmospheric pressure. If the water had an initial temperature of 25°C you would not expect to see much change in its appearance. However, if the initial temperature was 90°C, the same increase in temperature would lead to a dramatic change as a threshold was crossed and the water started to boil. In the second example, the sensitivity of the system is much greater because the water temperature is closer to boiling point. Once a threshold has been crossed, the system reaches a new equilibrium. Relating this back to the sand grain example, two scenarios could be considered. In the first, the particle is resting on a flat bed and is fully exposed to the flow. In the second, it is buried beneath a layer of gravel. The threshold for movement will be much lower for the exposed grain than for the buried grain, which cannot move unless the overlying gravel is removed.

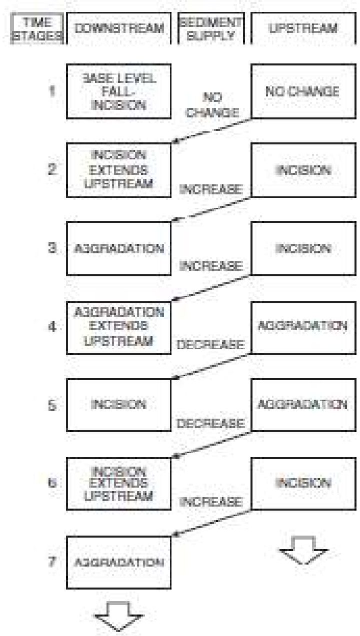


Figure 2.3 Schematic representation of complex response in a drainage system as a result of a fall in base level. See text for full explanation. After Summerfield (1991).

Complex response

The response of the fluvial system to change is often complex because of the many interrelationships that exist between the different components of the system. An example is the complex response of a tributary to a lowering in base level elevation at its outlet (Schumm, 1977). Here the main river, into which the tributary flows, has degraded, or lowered its channel elevation by erosion. This leads to a complex sequence of episodic erosion and deposition in the tributary as the system searches for a new equilibrium. A similar sequence of events can be observed from experimental simulations of laboratory drainage networks (Summerfield, 1991) and is illustrated schematically in Figure 2.3. This represents a sequence of events that occurs at the upstream end of a channel (boxes running down the right hand side of the figure) and at the downstream end (boxes running down the left hand side of the figure). The numbers on the left hand side (1 to 7) indicate sequential time stages. Starting at the top of the page and working down, we can see what is happening upstream and downstream at each time stage by reading across the page. Changes in sediment supply from upstream are also indicated by the arrows running from 'upstream' boxes to 'downstream' boxes.

● *Stage 1*. A fall in base level results in a local steepening of the channel gradient at the downstream end. This might exist as a sharp break in channel slope or as a steeper section of channel. Erosion is focussed at this steeper part, leading to a localised lowering in channel elevation. As a result of this lowering, the section of channel immediately upstream becomes steeper. The steep zone therefore migrates upstream over time. The initial response of the channel to the fall in base level is therefore incision (downcutting) at the downstream end. At this stage the upstream part of the channel is unaffected, so there is no change in the supply of sediment coming from upstream.

● *Stage 2*. Over time, incision is propagated upstream. As a result there is an increase in the amount of sediment coming from further upstream.

● *Stage 3*. This increase in sediment supply from upstream leads in turn to a build up of sediment in the channel at the downstream end. This process is called **aggradation** and results in a reduction in channel gradient. Meanwhile, the zone of incision continues to migrate further upstream.

● *Stage 4*. Aggradation is propagated upstream. Since there is now a net deposition of sediment in the upstream part of the channel, there is a decrease in the volume of sediment supplied from upstream.

● *Stage 5*. At the downstream end, less sediment means that there is additional energy for erosion and a new cycle of incision begins. Aggradation continues to occur at the upstream end.

● *Stage 6*. The wave of incision extends upstream, once again increasing the supply of sediment from upstream.

● *Stage 7*. The increased sediment supply leads to aggradation at the downstream end of the channel.

This example illustrates the way in which downstream changes can be propagated upstream, while the resultant upstream changes in turn control what happens downstream. In this way there is a complex cycle as periods of **episodic erosion** are interspersed with periods of deposition.

SCALE IN FLUVIALGEOMORPHOLOGY

Scale is an important consideration in fluvial geomorphology, with process–form interactions occurring over a huge range of space and time scales. At one end of this range is the long-term evolution of the landscape. At the other are small-scale processes, such as the setting in motion of an individual grain of sand resting on the bed of a channel. Space scales therefore encompass anything from a few millimetres to hundreds of kilometres. Relevant time scales stretch from a few seconds to hundreds of thousands of years or more. In order to understand how the fluvial system operates we can examine the relationships between processes and form in more detail at finer scales. This can be done by examining individual sub-systems, or sub-systems within sub-systems. When focusing in like this it is important to remember that these sub-systems are all part of an integrated whole and therefore cannot be considered in isolation from the rest of the system.

Space scales (spatial scales)

In studying the fluvial system, the scale of relevance varies according to the type of investigation. At the largest, drainage basin scale (Figure 2.4a), it is possible to see the form and characteristics of the drainage network and drainage basin topography. These reflect the cumulative effect of processes operating over long time scales, as well as past changes imposed by the external basin controls. At a smaller scale, the form of a reach of meandering channel can be examined in the context of drainage basin history and the influence of controlling variables at the channel scale, such as the supply of water and sediment from upstream (Figure 2.4b). The way in which the form and position of the channel has changed over time scales extending to thousands of years may be preserved as floodplain deposits, which can be used in reconstructing drainage basin history. Moving in to look at an individual meander bend (Figure 2.4c), process–form interactions can be observed at a smaller scale. These include flow hydraulics within the bend and associated sediment dynamics. Investigations of rates of bend migration or bank erosion processes are also carried out at this scale. Depositional channel units such as the point bar (seen on the inside of the meander bend in Figure 2.4c) are of interest to sedimentologists, providing evidence about the flows that formed them. At a finer scale still are individual ripples on the bar surface (Figure 2.4d) formed by the most recent high flow and, moving even closer, the internal arrangement of grains (Figure 2.4e). At the finest scale are individual grains of sediment (Figure 2.4f).

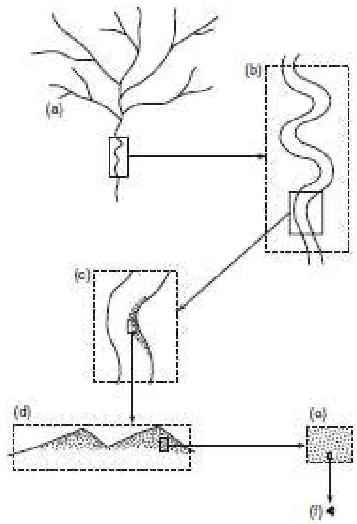


Figure 2.4 Idealised sketch showing the components of the fluvial system. (a) Channel network. (b) River reach. (c) Meander bend. (d) Bedforms. (e) Sedimentary structure. (f) Sediment grain. After Schumm (1988).

Time scales (temporal scales) and equilibrium

At smaller spatial scales, process–form interactions generally result in more rapid adjustments. At the largest scale, the long term evolution of channel networks occurs over time scales of hundreds of thousands of years or more, while the migration of individual meander bends can be observed over periods of years or decades, and small-scale flow-sediment interactions within minutes. The perspective of the historically oriented geomorphologist concerned with the large-scale, long-term evolution of landforms is therefore very different to that of the process geomorphologist or engineer who is interested in the operation of channel processes at much shorter time scales (Schumm, 1988). Historical studies show that the fluvial system follows an evolutionary sequence of development that is interrupted by major changes induced by the external basin controls. However, over the much shorter time periods involved in the field measurement of processes, there may be little or no significant change in fluvial landforms. This might not matter too much if flow–sediment interactions at very small scales are of interest, although basin history certainly does have an influence at the reach scale, since channel form has been shaped by past changes in flow and sediment supply.

The precise definition of equilibrium is also time dependent. Equilibrium refers to a state of balance within a system, or sub-system. Negative feedback mechanisms help to maintain the system in an equilibrium state, buffering the effect of changes in the external variables. However, different types of equilibrium may exist at different time scales. These were defined by Schumm (1977) with reference to changes in the elevation of the bed of a river channel above sea level. If you were to observe a short section of river channel over a period of a few hours you would not see any change in its form (unless there happened to be a flood), although you might see some sediment transport. Over this short time period the channel is said to be in a state of **static equilibrium** (Figure 2.5a). The same river, observed over a longer time scale of a decade, would show some changes. During this time, floods of various sizes pass through the channel, scouring the bed. In the intervening periods, deposition builds up the channel bed again. As a result of these cycles of scour and fill, the elevation of the channel bed fluctuates around a constant average value (Figure 2.5b) and **steady state equilibrium** exists.

Over longer time scales, from thousands to hundreds of thousands of years or more, erosion gradually lowers the landscape. At these time scales, the channel elevation fluctuates around a changing average condition, the underlying trend being a reduction in channel elevation. This is called **dynamic equilibrium** and is illustrated in Figure 2.5c. As you know, the influence of the external basin controls cannot be ignored. Changes in any of these variables can lead to positive feedbacks within the system and a shift to a new equilibrium state. For example, in tectonically active regions, the section of channel might be elevated by localised uplift. Such episodes of change occur over much shorter time scales than the gradual evolution of the landscape, resulting in abrupt transitions. This type of equilibrium delights in the term **dynamic metastable equilibrium** (Figure 2.5d).

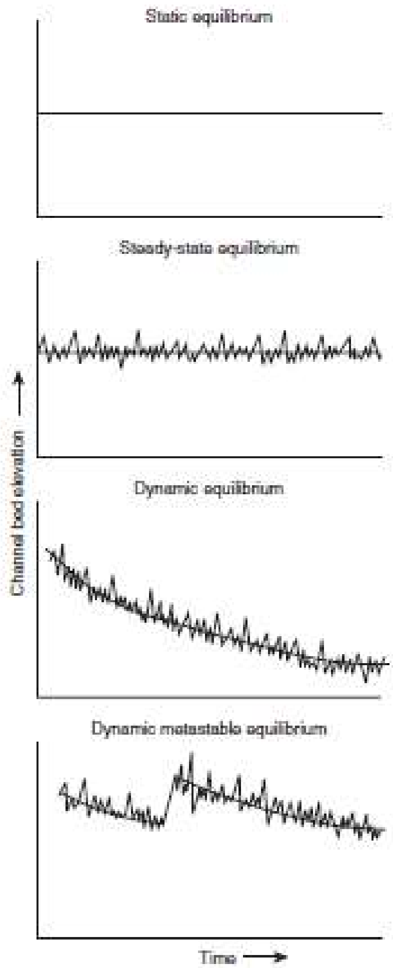


Figure 2.5 Types of equilibrium observed over different time scales. Adapted from Schumm (1977).

UNIT -4: CHANNEL INITIATION: THEORY OF OVERLAND FLOW, THEORY OF SUB-SURFACE FLOW

FLOW GENERATION

Hydrological pathways

Inputs of water to a drainage basin are the various forms of **precipitation** that fall over its area. These include rain, snow, sleet, hail and dew. If you look at a drainage basin on a map, you will see that river channels cover only a very small part of the total area. This means that most of the water reaching the ground surface must find its way from the hillslopes and into the channel network. A number of pathways are involved (Figure 3.1), and a given 'parcel' of water arriving at a stream channel may have taken any number of them. Not all the incoming precipitation actually makes it to the basin outlet, since a certain percentage is evaporated back to the atmosphere. Water that falls on the leaves of vegetation and artificial structures like buildings is **intercepted**. Some of this water does fall to the ground, as anyone who has sheltered under a tree will know, but much of it is evaporated (water resource managers refer to this output as 'interception loss'). The amount of interception that occurs is dependent on such factors as the extent and type of vegetation (leaf size, structure, density and the arrangement of foliage), wind speed and rainfall intensity (Jones, 1997). Evaporation also takes place from the surface layers of the soil, and from lakes and wetland areas. A related process is **transpiration**, whereby plants take up water through their roots and evaporate it through pores, called stomata, on the underside of their leaves. This means that water can be 'lost' to the atmosphere from some depth below the surface. Because evaporation and transpiration are difficult to monitor separately, they are usually considered together as the combined process of **evapotranspiration**. Water reaching the soil surface may either enter the soil by a process called **infiltration** or remain at the surface, moving down-slope as **overland flow** (labelled 1 and 4 in Figure 3.1). Infiltrated water either travels through the soil, parallel to the surface, as **through flow** (2), or slowly percolates downwards to the saturated zone, travelling as **groundwater flow** (3). Rates of water movement are fastest for overland flow (typically between 50 m and 500 m per hour), between 0.005 m and 0.3 m per hour for throughflow and 0.005 m to 1.5 m per *day* for groundwater flow (Ward and Robinson, 1990). How fast a river rises after rainfall is very much dependent on the relative proportion of water taking faster (surface and near-surface) pathways, and slower (subsurface) pathways. An important control on this is the rate at which water infiltrates into the soil, which is determined by the **infiltration capacity**. This is defined as the maximum rate at which water enters the soil when it is in a given condition (Horton, 1933). This is an important definition: if the rainfall intensity exceeds the infiltration capacity, the soil cannot absorb all the water and there is excess water, which ponds at the surface. This moves down slope as overland flow, more specifically, **Hortonian overland flow** (1). As Horton's definition implies, soil properties control the infiltration capacity; these include soil permeability, the presence of vegetation and plant roots and how much water is already in the soil.

Following a dry period, the infiltration capacity is highest at the start of a storm, rapidly decreasing as rainfall continues, until a constant infiltration rate is reached (Horton, 1933).

A second type of overland flow, **saturation overland flow** (4), is generated when the soil is totally saturated and therefore cannot take any more water (Kirkby and Chorley, 1967). Where the water table is relatively shallow, for example in valley bottoms, rainfall can cause it to rise to the ground surface. Where this occurs, a saturated area forms around the channel, increasing in extent as the storm progresses. The saturated area acts as an extension to the channel network, meaning that a significant volume of water is transferred in a short time. These saturated areas are known as **variable source areas** or **dynamic contributing areas** and are highly significant in humid environments, where this is the main way in which storm run-off is generated (Hewlett and Hibbert, 1967; Dunne and Black, 1970). Hortonian overland flow is rarely observed in humid environments, unless the surface has a low infiltration rate, for example where there are outcrops of bare rock or artificially paved surfaces. However, in dryland environments, a combination of factors mean that Hortonian overland flow is the dominant mechanism. Rainfall, when it occurs, typically has a high intensity and exceeds the low infiltration capacity of sparsely vegetated soils (Dunne, 1978). In addition, dryland soils often develop a crust at the surface, which further reduces the infiltration capacity. Water that infiltrates the soil may travel at various depths below the surface. While it is fairly obvious that water should move downwards under the force of gravity, it might seem counter-intuitive that it also flows through the soil as throughflow. This happens because a preferential flow path is set up. Soil permeability decreases with depth, meaning that the downward movement of water is slowed. During rainfall, this leads to a backing up effect in the more permeable surface layers. This has been likened to the flow of water down a thatched roof: it is easier for the water to move parallel to the slope of the roof, along the stems of the straw, than to move vertically downwards through it (Ward, 1984; Zaslavsky and Sinai, 1981). Where they exist, **soil pipes** provide a very rapid throughflow mechanism, and rates of flow can be comparable to those in surface channels. Soil pipes are hydraulically formed conduits that can be up to a metre or more in diameter. They are found in a wide range of environments and are sometimes several hundreds of metres in length (Jones, 1997). As such, they can act as an extension to the channel network, allowing the drainage basin to respond rapidly to precipitation inputs (Jones, 1979). The schematic diagram in Figure 3.1 represents the headwaters of a humid zone river, where the channel intersects the water table surface (it should be noted that aquifers are not always present, for example aquifer development is extremely limited in headwater areas that are underlain by impermeable rocks and characterised by thin soils). In this example, groundwater contributions are made to the channel flow from the underlying aquifer and the channel is called a **gaining stream**. Even during rainless periods, flow will be maintained, as long as the level of the water table does not fall below that of the channel bed. A rather different situation exists for many dryland channels, where the water table may be several metres, or tens of metres, below the surface. In this case, the direction of flow is reversed, as water is lost through the bed and banks of the channel, percolating downwards to recharge the aquifer. This is termed a **losing stream** and, although not exclusive to drylands, many examples are found in these environments. It is not unusual for a river to be gaining and losing flow to groundwater along different parts of its course, while seasonal fluctuations in water table levels can mean that losing streams become gaining streams for part of the year. The Euphrates in Iraq provides a good example, with most of the flow being generated in the headwaters in northern Iraq, Turkey and Syria. Further downstream, at the Hit gauging station (150 km west of Baghdad) the river loses flow to groundwater for much of the year (Wilson, 1990). The loss of flow from a channel, due to downward percolation and high evaporation rates, is referred to as **transmission loss**.

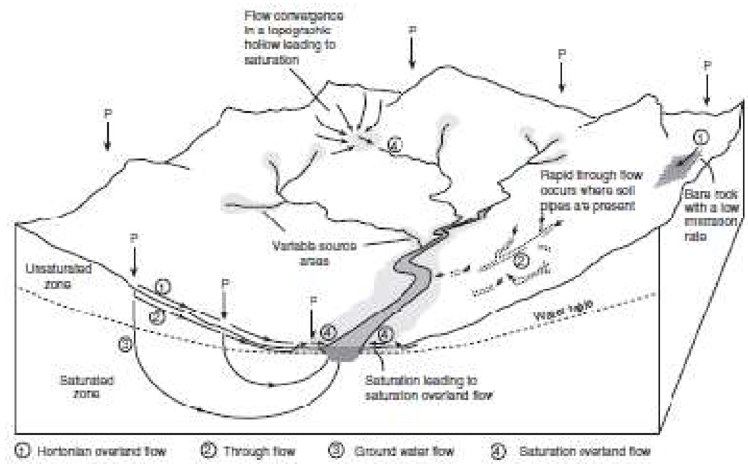


Figure 3.1 Surface and subsurface hydrological pathways.

UNIT – 5: FLOW REGIME: HYDROLOGICAL PATHWAYS, MEASUREMENT OF STREAM FLOW AND ANNUAL FLOW REGIMES; CLASSIFICATION OF NATURAL STREAMS BY D.L. ROSGEN

The flow in river channels exerts hydraulic forces on the boundary (bed and banks). An important balance exists between the erosive force of the flow (driving force) and the resistance of the boundary to erosion (resisting force). This determines the ability of a river to adjust and modify the morphology of its channel. One of the main factors influencing the erosive power of a given flow is its discharge: the volume of flow passing through a given cross-section in a given time. Discharge varies both spatially and temporally in natural river channels, changing in a downstream direction and fluctuating over time in response to inputs of precipitation. Characteristics of the flow regime of a river include seasonal variations in discharge, the size and frequency of floods and frequency and duration of droughts. The characteristics of the flow regime are determined not only by the climate but also by the physical and land use characteristics of the drainage basin.

FLOW GENERATION

Hydrological pathways

Inputs of water to a drainage basin are the various forms of **precipitation** that fall over its area. These include rain, snow, sleet, hail and dew. If you look at a drainage basin on a map, you will see that river channels cover only a very small part of the total area. This means that most of the water reaching the ground surface must find its way from the hillslopes and into the channel network. A number of pathways are involved (Figure 3.1), and a given 'parcel' of water arriving at a stream channel may have taken any number of them. Not all the incoming precipitation actually makes it to the basin outlet, since a certain percentage is evaporated back to the atmosphere. Water that falls on the leaves of vegetation and artificial structures like buildings is **intercepted**.

Some of this water does fall to the ground, as anyone who has sheltered under a tree will know, but much of it is evaporated (water resource managers refer to this output as 'interception loss'). The amount of interception that occurs is dependent on such factors as the extent and type ofvegetation (leaf size, structure, density and the arrangement of foliage), wind speed and rainfall intensity (Jones, 1997). Evaporation also takes place from the surface layers of the soil, and from lakes and wetland areas. A related process is **transpiration**, whereby plants take up water through their roots and evaporate it through pores, called stomata, on the underside of their leaves. This means that water can be 'lost' to the atmosphere from some depth below the surface. Because evaporation and transpiration are difficult to monitor separately, they are usually considered together as the combined process of **evapotranspiration**. Water reaching the soil surface may either enter the soil by a process called **infiltration** or remain at the surface, moving down-slope as **overland flow** (labelled 1 and 4 in Figure 3.1). Infiltrated water either travels through the soil, parallel to the surface, as **throughflow** (2), or slowly percolates downwards to the saturated zone, travelling as **groundwater**

flow (3). Rates of water movement are fastest for overland flow (typically between 50 m and 500 m per hour), between 0.005 m and 0.3 m per hour for throughflow and 0.005 m to 1.5 m per *day* for groundwater flow (Ward and Robinson, 1990). How fast a river rises after rainfall is very much dependent on the relative proportion of water taking faster (surface and near-surface) pathways, and slower (subsurface) pathways. An important control on this is the rate at which water infiltrates into the soil, which is determined by the **infiltration capacity**. This is defined as the maximum rate at which water enters the soil when it is in a given condition (Horton, 1933). This is an important definition: if the rainfall intensity exceeds the infiltration capacity, the soil cannot absorb all the water and there is excess water, which ponds at the surface. This moves down slope as overland flow, more specifically, **Hortonian overland flow** (1). As Horton's definition implies, soil properties control the infiltration capacity; these include soil permeability, the presence of vegetation and plant roots and how much water is already in the soil. Following a dry period, the infiltration capacity is highest at the start of a storm, rapidly decreasing as rainfall continues, until a constant infiltration rate is reached (Horton, 1933). A second type of overland flow, **saturation overland flow** (4), is generated when the soil is totally saturated and therefore cannot take any more water (Kirkby and Chorley, 1967). Where the water table is relatively shallow, for example in valley bottoms, rainfall can cause it to rise to the ground surface. Where this occurs, a saturated area forms around the channel, increasing in extent as the storm progresses. The saturated area acts as an extension to the channel network, meaning that a significant volume of water is transferred in a short time. These saturated areas are known as **variable source areas** or **dynamic contributing areas** and are highly significant in humid environments, where this is the main way in which storm run-off is generated (Hewlett and Hibbert, 1967; Dunne and Black, 1970). Hortonian overland flow is rarely observed in humid environments, unless the surface has a low infiltration rate, for example where there are outcrops of bare rock or artificially paved surfaces. However, in dryland environments, a combination of factors mean that Hortonian overland flow is the dominant mechanism. Rainfall, when it occurs, typically has a high intensity and exceeds the low infiltration capacity of sparsely vegetated soils (Dunne, 1978). In addition, dryland soils often develop a crust at the surface, which further reduces the infiltration capacity.

Water that infiltrates the soil may travel at various depths below the surface. While it is fairly obvious that water should move downwards under the force of gravity, it might seem counter-intuitive that it also flows through the soil as throughflow. This happens because a preferential flow path is set up. Soil permeability decreases with depth, meaning that the downward movement of water is slowed. During rainfall, this leads to a backing up effect in the more permeable surface layers. This has been likened to the flow of water down a thatched roof: it is easier for the water to move parallel to the slope of the roof, along the stems of the straw, than to move vertically downwards through it (Ward, 1984; Zaslavsky and Sinai, 1981). Where they exist, **soil pipes** provide a very rapid throughflow mechanism, and rates of flow can be comparable to those in surface channels. Soil pipes are hydraulically formed conduits that can be up to a metre or more in diameter. They are found in a wide range of environments and are sometimes several hundreds of metres in length (Jones, 1997). As such, they can act as an extension to the channel network, allowing the drainage basin to respond rapidly to precipitation inputs (Jones, 1979).

The schematic diagram in Figure 3.1 represents the headwaters of a humid zone river, where the channel intersects the water table surface (it should be noted that aquifers are not always present, for example aquifer development is extremely limited in headwater areas that are underlain by impermeable rocks and characterised by thin soils). In this example, groundwater contributions are made to the channel flow from the underlying aquifer and the channel is called a **gaining stream**. Even during rainless periods, flow will be maintained, as long as the level of the water table does not fall below that of the channel bed. A rather different situation exists for many dryland channels, where the water table may be several metres, or tens of metres, below the surface. In this case, the direction of flow is reversed, as water is lost through the bed and banks of the channel, percolating downwards to recharge the aquifer. This is termed a **losing stream** and, although not exclusive to drylands, many examples are found in these environments. It is not unusual for a river to be gaining and losing flow to groundwater along different parts of its course, while seasonal fluctuations in water table levels can mean that losing streams become gaining streams for part of the year. The Euphrates in Iraq provides a good example, with most of the flow being generated in the headwaters in northern Iraq, Turkey and Syria. Further downstream, at the Hit gauging station (150 km west of Baghdad) the river loses flow to groundwater for much of the year (Wilson, 1990). The loss of flow from a channel, due to downward percolation and high evaporation rates, is referred to as **transmission loss**.

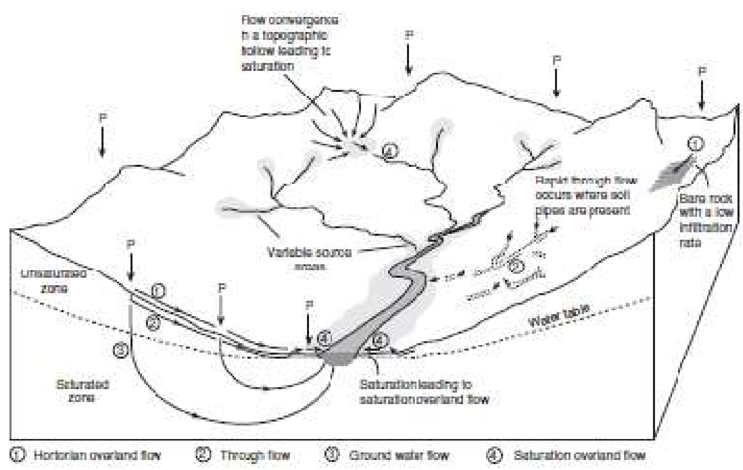


Figure 3.1 Surface and subsurface hydrological pathways.

The storm hydrograph and drainage basin response

Flow discharge (also known as ***Q***) is the volume of water passing through a given channel cross-section in a given time. The units of discharge most commonly used are cubic metres per second (m3 s−1), known as 'cumecs', although for very small flows litres per second may be used. In the United States cubic feet per second, or 'cusecs', are used instead of cumecs. Box 3.1 explains how discharge is monitored.

THE MEASUREMENT OF STREAMFLOW

The channel discharge is the volume of water flowing through a given channel cross-section in a given time. A number of different methods have been developed to measure discharge. These can be grouped into **instantaneous measurements**, where discharge is measured at a particular point in time, and **continuous measurements** for a record of discharge variations through time.

Box 3.1

THE MEASUREMENT OF STREAMFLOW—CONT'D

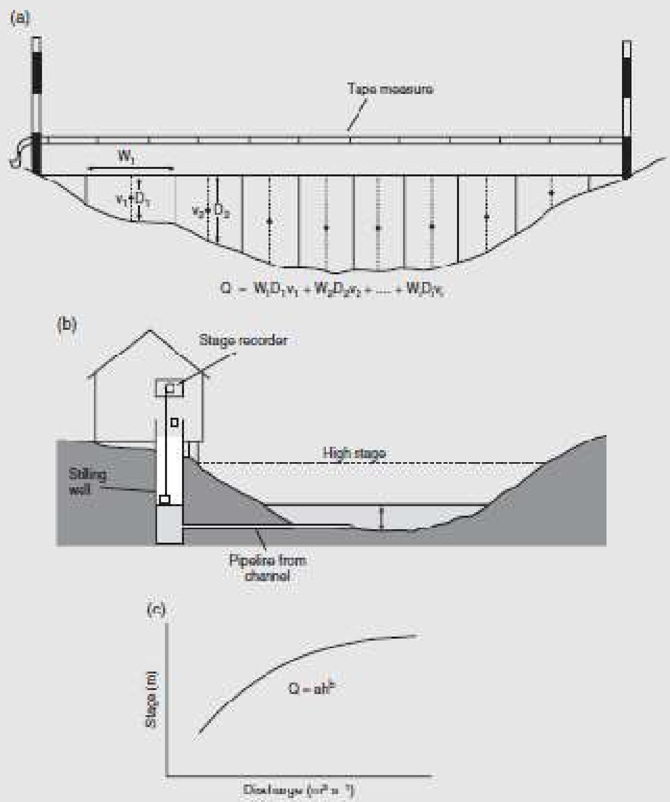


Figure 1 (a) Measuring discharge using the velocity-area method. Q = discharge, D = depth, w = width and v = velocity. (b) Gauging station at a rated section. (c) Rating curve. Q = discharge, h = stage, a and b are coefficients.

(Continued)

The velocity–area method

Discharge is measured in cubic metres per second (m3s−1). It increases with the area of the channel cross section and with the velocity of flow. Discharge can be calculated for a given channel cross-section by measuring its cross-sectional area and the mean flow velocity:

Q □□A □□v

where Q □□discharge, A □□cross-sectional area and v □ mean flow velocity.

Figure 1(a) illustrates the method used. The first stage is to stretch a tape measure across the width of the channel. The channel is then divided into a number of sub-sections. This varies according to the width of the river, but discharge is usually gauged at twenty or more subsections. Ideally, the discharge flowing through each sub-section should be similar, so these are more closely spaced where the flow is deeper and faster. Velocity measurements are then made using a flow meter. A commonly used design is a propeller mounted on a rod, which is lowered into the flow. The flow velocity is directly proportional to the rate at which the propeller is turned by the flow. A digital readout shows the velocity in m s−1 or, if you are not so lucky, the number of rotations per minute. (This can then be converted to the equivalent velocity using a simple formula.) One velocity measurement is made for each of the sub-sections, at the points indicated in Figure 1(a). Since flow velocity increases from zero at the channel bed to a maximum near the water surface, a representative average flow needs to be measured. It can be shown that the flow velocity at a height of 0.4*d* above the bed is representative of the average flow velocity, where *d* is the total depth of flow. Velocity measurements should therefore be made at a distance below the water surface that is 0.6 of the total depth (i.e. 0.4 of the depth from the bed). Thus if the depth of flow was 1 m, you would measure the velocity at a depth of 0.6 m below the surface (or 0.4 m above the bed).

The width and depth of each sub-section are measured as shown in the diagram. The discharge flowing through each sub-section can be calculated by multiplying the sub-section width (*w*), depth (*d*) and velocity (*v*). In order to calculate the discharge for the whole cross-section, the discharges for each of the subsections are added together. The discharge flowing through the 'left over' triangles adjacent to each bank is usually assumed to be negligible and is not included in this calculation.

This method is not very suitable for steep, turbulent, rocky streams where accurate current meter measurements are hard to obtain. A more appropriate technique in this case is to use **dilution gauging**. A chemical 'tracer' substance such as salt or dye is released into the flow, either as a single large 'gulp' or 'slug', or by continuously injecting it into the flow. Changes in concentration are monitored further downstream. Since the amount of dilution increases with discharge, it is possible to relate the change in concentration to stream discharge. **Continuous streamflow measurement** Although laborious, the velocity-area method only gives you one measurement of discharge at a particular point in time. In order to plot hydrographs like the ones in Figure 3.2, it is necessary to have a continuous record of flow. Discharge is difficult to measure directly. However, it is related to **stage**, the water level or height, which is much easier to record. Figure 1(b) shows a gauging

THE MEASUREMENT OF STREAMFLOW—CONT'D

station situated on a natural cross-section. Rather than measuring the water level in the river itself, a pipeline leads to a **stilling well** which damps out the effects of surface waves and turbulence. The level of water in the stilling well is monitored using a **stage recorder**. Traditionally this would have consisted of a float attached to a counterweight via a pulley (shown in Figure 1b). As the water level rises and falls the float moves up and down, turning the pulley. This is attached to a pen, which draws a trace on a chart mounted on a rotating drum. The disadvantage is that the charts then have to be read manually. Digital measurements of water level can be made by sending an infrared signal down the stilling well. This is reflected by the water surface back to a receiver, with the time delay measured. Stage is converted to discharge using a **rating curve** (Figure 1c). Using this, the discharge corresponding to a given stage can either be read off from the graph or calculated using the rating equation. This is of the form $Q = ah^b$, where $Q =$ discharge, $h =$ stage and a and b are coefficients, which describe the unique relationship between stage and discharge for the cross-section. This relationship is affected by the shape of the crosssection. At lower flows, small increases in discharge result in relatively large increases in stage because the channel is narrow near its bed. As the flow increases, a greater increase in discharge is needed to produce the same increase in depth, because the channel widens above the bed. Therefore a wide, shallow channel will have a different relationship (relatively small increases in stage with rising discharge) to a narrow, deep channel (relatively large increases in stage with rising discharge).

The rating curve is derived from discharge measurements made at different stages. Measurements are difficult to make at high flows, so most rating curves tend to be less reliable at high flows. Another problem is that if a large flood alters the shape of the cross-section, the rating curve has to be re-calibrated. This can be avoided using a gauging structure such as a **flume** or **weir**. These have a regular cross-sectional area and the flow velocity is controlled by the structure. The rating curve can be derived using a combination of measurements and hydraulic theory. The top graph in Figure 3.2 (solid line) shows an **annual hydrograph** of daily flows, which might be observed for a river in the temperate zone over the course of a year. The lower graph is a single **storm hydrograph**, which shows the response of the drainage basin to one precipitation event. During a particular rainfall event, there is a delay between the onset of rainfall and the time at which the discharge starts to increase. The initial increase is due to water falling directly into the channel and close to it, though as the storm progresses, water travelling from greater distances reaches the channel. Water taking the fastest pathways – overland flow, shallow throughflow and pipe flow – is rapidly transferred to the channel, contributing to the **quick flow** component of the storm hydrograph. **Base flow** contributions come from water taking the slower subsurface routes, taking much longer to reach the channel. This means that water continues to enter the channel as base flow for some time after rainfall has ceased, keeping rivers flowing during dry periods. Glaciers, lakes, reservoirs and wetlands also contribute to base flow. The relative proportions of quick flow and base flow determine the size of the hydrograph peak and the time delay, or **lag time**, between peak rainfall and peak flow. For instance, where quick flow dominates, lag times are relatively short and peak flows relatively high. Rivers dominated by base flow respond more slowly and the peak flow is lower. The two lines shown in Figure 3.2(a) represent two temperate zone rivers, identical in every respect apart from the underlying geology: basin 1 is underlain by impermeable rocks and basin 2 by permeable rock. From the graphs you will see that basin 1 has a flashy response, with marked storm peaks, whereas basin 2 has a more damped response.

This is because most of the precipitation falling over basin 2 infiltrates and takes slower subsurface pathways to the stream channel. There is also a marked difference in the summer low flows, with the greater base flow component for basin 2, which sustains higher summer flows. The hydrological response of a river to discrete inputs of precipitation through time, as indicated by the shape of its hydrograph, is determined by the drainage basin characteristics and climatic factors shown in Table 3.1.s

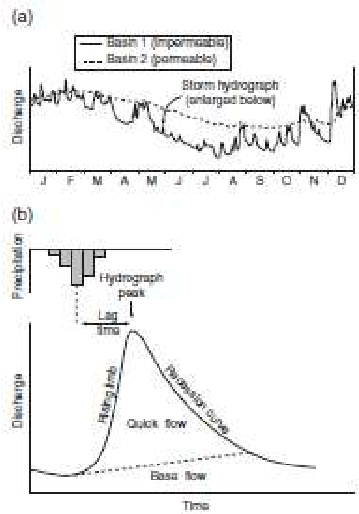


Figure 3.2 (a) Typical annual hydrograph of a temperate zone river whose drainage basin is underlain by an impermeable geology (solid line). Also shown is the annual hydrograph of a basin with a more permeable geology (dotted line). (b) Characteristics of a storm hydrograph.

ANNUAL FLOW REGIMES

The **annual flow regime** of a river describes the seasonal variations in flow that are observed during an 'average' year. As you might expect, this is influenced by the seasonal distribution of rainfall, and the balance between rainfall and evaporation at different times of year. For example, some tropical rivers experience a marked wet and dry season, drying up completely for part of the year and carrying high flows during the wet season. Climate also has an important influence on the type and density of vegetation, soils and land use, all of which act as controls on the processes of runoff generation (Table 3.1). Several climate characteristics are important in determining the flow regime. These include whether it is humid or arid, if it is predominantly warm or cold, the annual range of temperatures, and whether precipitation is seasonal or occurs all year round. At high latitudes and in some mountain environments, the timing and length of glacial ablation and snowmelt is a dominant factor. Figure 3.3 shows a selection of typical flow regimes, which characterise different climatic zones. These come from a classification scheme developed by Beckinsale (1969) from an existing climate classification. The different regimes are categorized using a system of two letters. The first letter relates to the mean annual precipitation and annual temperature range:

A: Warm, moist tropical climates, where the mean temperature exceeds 18°C for all months of the year.

B: Dry climates, where rates of potential evaporation1 exceed annual precipitation.

C: Warm moist temperate climates.

D: Seasonally cold climates with snowfall, where the mean temperature is less than –3°C during the coldest month. The second letter indicates the seasonal distribution of precipitation:

F: Appreciable rainfall all year round.

W: Marked winter low flow.

S: Marked summer low flow. For example, the regime of the Pendari River (Figure 3.3A) is influenced by a tropical climate with a marked winter low flow, and would be classified as AW.

Table 3.1 Factors affecting hydrological response of a basin

Soils and geology

● Soil type and thickness. Soil texture (relative proportion of sand, silt and clay particles) affects infiltration rates. Sandy soils have high permeability whereas clay soils do not. In arid areas a crust can form on the soil surface, decreasing the permeability. Soil thickness affects how much water the soil can absorb.

● Geology. Drainage basins underlain by a permeable geology tend to have a slower response to precipitation, although the flow is sustained for a longer time during dry periods. Drainage basins underlain by impermeable materials have a faster, or more 'flashy' response.

Vegetation and land use

● Vegetation type and density. Vegetation reduces the impact of raindrops and allows a more 'open' soil structure, meaning that infiltration rates are higher. Vegetation also affects interception rates and evapotranspiration losses from the basin.

● Urban areas. Depends on the proportion of the drainage basin that is urbanised. Large areas of paved surfaces, drains and culverts rapidly transmit water to river channels, leading to an increase in peak flow and a shorter lag time.

● Grazing and cultivation. When deforestation occurs, rates of overland flow tend to increase. Heavy machinery and trampling by animals compact the soil, reducing permeability, although ploughing can increase infiltration rates. Flow may be concentrated in plough furrows that run up and down the slope.

● Land drainage. The installation of field drains allows rapid transfer of runoff into the nearest stream channel.

Physiographic characteristics

● Drainage basin size and shape. In larger basins the travel times are longer, as flow has to travel greater distances to reach the outlet. The total volume of runoff increases with the drainage area. Elongated drainage basins have a response that is initially more rapid but with a lower, more gentle peak.

● Drainage density. Where the density of stream channels is high, the average distance over which water has to travel to reach the channel network is reduced, leading to a more rapid response.
● Drainage basin topography. Travel times are increased over steep slopes. In upland areas, steep slopes are often associated with thin soils and the response tends to be flashy. Rainfall may be affected by altitude and aspect with respect to storm tracks.

Channel characteristics

● Channel and floodplain resistance. The velocity of flow in river channels is affected by the roughness of the bed and banks and the shape of the channel. Overbank flows are slowed by the roughness of the floodplain surface.
● Floodplain storage. When channel capacity is exceeded, water spills out onto the surrounding floodplain, where it is stored until the floodwaters recede. If floodplain storage is limited, a greater volume of water travels downstream.
● Conveyance losses. In dryland environments the channel may lose flow due to high rates of evaporation and 'leakage' by exfiltration through the channel boundary.

Meteorological factors

● Antecedent conditions: The conditions in the drainage basin prior to the onset of precipitation. Where recent or prolonged previous rainfall has occurred, the soil may be near saturation, meaning that a relatively small input of rainfall could lead to a rapid runoff response. Where snow is lying on the ground, subsequent rainfall can cause it to melt, which may lead to flooding downstream.
● Rainfall intensity: Rainfall intensity is expressed in millimetres per hour (mm h^{-1}). The more intense the rainfall, the more likely it is that the infiltration capacity of the soil will be exceeded.
● Rainfall duration: This is the period of time over which a given rainfall event takes place. As the storm progresses, runoff contributing areas at greater and greater distances from the channel network become active. The channel network may also extend upstream as normally dry channels start to carry flow.

The Arno (Figure 3.3C) has a warm temperate rainy climate, with summer low flows, and is classified as CS. On first looking at Figure 3.3, it might seem strange that the graph for dry climates (B) has the biggest peak. Bear in mind that these graphs indicate annual flow variability, rather than actual monthly discharges. The value for each month is the ratio of the monthly mean flow to the overall (annual) mean, which is shown by the dotted line on each graph. Regimes associated with **tropical, rainy climates** (Figure 3.3A) are affected by seasonal shifts in the inter-tropical convergence zone.

Near the equator, runoff occurs year-round, with peaks at the equinoxes (Lobaye River). Further north and south, there are marked wet and dry seasons (Pendjari River). The annual runoff for rivers in **dry climates** (Figure 3.3B) is low, but extremely variable. For much of the time there is little or no flow, but extreme floods can also occur. Flood pulses on the Cooper Creek are highly erratic and do not occur every year, since they are associated with El Nino disturbances to the monsoon. In contrast, there is much less variability for rivers with **temperate rainy climates** (Figure 3.3C). These have higher flows in winter and relatively low flows in summer, a pattern which is accentuated in Mediterranean climates (Arno). Snowmelt peaks are seen for rivers with **seasonally cold, snowy climates** (Figure 3.3D). The timing of the peak is dependent on altitude, latitude and seasonal patterns of rainfall. In **high mountain environments** (Figure 3.3H), class HN denotes drainage basins with a nival regime (where snow patches exist) and HG those with a glacial regime. H stands for *Höhenklima*, which translates from the German as 'highland climate'.

For any drainage basin, inter-annual variations will occur between wet and dry years, so the flow regime is described using long-term averages. Even within a particular climate zone, the flow regime can differ markedly between drainage basins. This is because the physical characteristics of a drainage basin also play an important role, most notably the underlying geology and soil characteristics. These determine how much storage is available within the drainage basin in natural reservoirs, such as groundwater, lakes and wetlands.

Downstream variations in discharge

As well as varying through time, discharge also changes along the course of a river. At any location, channel form is dependent on the discharge and supply of sediment from upstream. In most cases, discharge increases downstream as the area of the drainage basin increases and tributaries join the main channel. There is also a general increase in the size of the channel, with discharge acting as a control on the gross dimensions (Knighton, 1998). The quantitative description and understanding of the nature of these downstream changes have been the focus of much research and is explored further in Chapter 8. Although there is a *general* downstream increase in channel dimensions, local influences lead to considerable variation, even over short distances. Downstream changes in dryland channels can be very pronounced (Tooth, 2000). For example, infrequent flooding occurs along the ephemeral streams draining the Barrier Range in arid western New South Wales, Australia. Away from the uplands, high transmission losses lead to a rapid downstream reduction in discharge and channel size (Dunkerley, 1992). Downstream reductions in discharge and cross-sectional area are also observed in the piedmont and lowland zones of rivers draining the northern plains of arid central Australia (Tooth, 2000). There is an eventual termination of channel flow and bedload transport. However, during large floods, flows continue out across extensive unchannelled surfaces called 'floodouts' (Tooth, 2000). Tooth highlights the complex interactions between discharge, sediment transport, channel slope, tributary inflows, bank sediments and vegetation. These give rise to considerable variations in the downstream channel changes observed for dryland rivers.

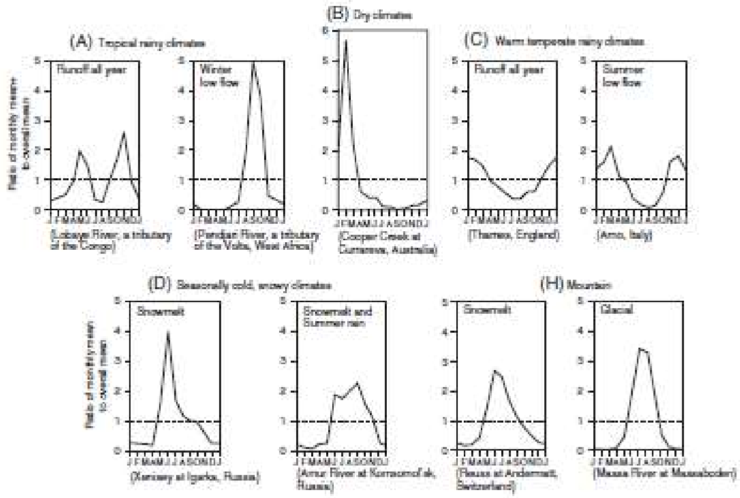


Figure 3.3 Flow regimes for selected rivers. (A) Tropical rainy climates: Lobaye River (a tributary of the Congo) and Pendjari River (a tributary of the Volta in West Africa). (B) Dry climates: Cooper Creek at Currareva, Australia. (C) Warm, temperate, rainy climates: Thames, England, and Arno, Italy. (D) Seasonally cold, snowy climates: Xenisey at Igarka and Amur River at Komsomol'sk, both in Russia. (H) Mountain climates: Reuse at Andermatt and Mass River at Massaboden, both in Switzerland. Source: Adapted from Beckinsale (1969); (B) after Knighton and Nanson (1997).

UNIT – 6: DRAINAGE EVOLUTION: PARALLEL DEVELOPMENT AND COALESCENCE, HEADWARD EXTENSION AND BRANCHING, LATERAL EXPANSION

Order analysis—The first step in drainage basin analysis is order designation, following a system only slightly modified from ***Horton*** [1945, p. 2 8 1 - 282] (Fig. 2). Assuming that the channel-network map includes all intermittent and permanent flow lines located in clearly defined valleys, the smallest finger-tip tributaries are designated Order 1. Where two first-order channels join, a channel segment of Order 2 is formed; where two of Order 2 join, a segment of Order 3 is formed; and so forth. The trunk stream through which all discharge of water and sediment passes is therefore the stream am segment of highest order. Any usefulness which the stream order system m a y h a v e depends upon t h e premise t h a t on t he average, if a sufficiently large sample is t r e a t e d, order number is directly proportional to relative watershed dimensions, channel size, and s t r e am discharge a t t h a t place in t h e system. Also, because order number is dimensionless, two drainage basins differing greatly in linear scale can be equated or compared w i t h respect t o corresponding p o i n t s in their geometry through use of order number. T h e first step in drainage-network analysis is t h e counting of s t r e am segments of each order. T h i s is followed b y analysis of the way in which n u m b e r s of s t r e am segments change with increasing order.

Bifurcation ratio—Horton's [1945, p . 291] law of s t r e am n u m b e r s s t a t e s t h a t t h e numbers of stream segments of each order f o rm an inverse geometric sequence with order number. This is generally verified by a c c u m u l a t e d d a t a ***[Strahler,*** 1952, p. 1137; ***Schumm,*** 1956, p .603] and is conveniently t r e a t e d as shown in F i g u r e 3. A regression of l o g a r i t hm of number of streams of each order (ordinate) on s t r e am order (abscissa) generally yields a s t r a i g h t - l i n e p l o t w i t h very little scatter ***[Maxwell,*** 1955]. E v e n t h o u g h t h e function relating these v a r i a b l e s is defined only for integer values of t h e i n d e p e n d e n t v a r i a b l e , a regression line is fitted; t h e slope of t h e line, or regression coefficient ***h*** is used. T h e a n t i - l o g a r i t hm of ***b*** is equivalent to H o r t o n ' s bifurcation r a t i o ***n*** and in this case has t h e v a l u e of 3.52. T h i s m e a n s t h a t on the average t h e r e a r e t h r e e a n d one-half times as m a n y streams of one order as of t h e n e x t higher order. One m i g h t t h i n k t h a t t h e bifurcation ratio would c o n s t i t u t e a useful dimensionless number for expressing t h e form of a d r a i n a g e system. Actually t h e number is highly s t a b l e and shows a small range of v a r i a t i o n from region t o region or environment to environment, except where powerful geologic controls d o m i n a t e . ***Coates*** [1956, Table 3j found bifurcation r a t i o s of first-order to second order streams to r a n g e from 4.0 to 5 . 1 ; r a t i o s of second-order to t h i r d - order streams t o range from 2.8 to 4.9. T h e s e v a l u e s differ l i t t l e f r om ***Strahler's*** [1952, p . 1134].

Frequency distribution of stream lengths—Length of stream channel is a dimensional p r o p e r t y which a n be used to reveal t h e scale of u n i t s comprising the drainage network. One method of length analysis is the measurement of length of each segment of channel of a given stream order. For a given watershed these lengths can be studied by frequency distribution analysis ***[Schumm,*** 1956, p. 607]. Stream lengths are strongly skewed right, but this may be largely corrected by use of logarithm of length. Arithmetic mean, estimated population variance, and standard deviation serve as standards of description whereby different drainage nets can be compared and their differences tested statistically ***[Strahler,*** 1954b].

Relation of stream length to stream order—Still another means of evaluating length relationships in a drainage network is to relate stream length to stream order. A regression of logarithm of total stream length for each order on logarithm of order may be plotted (Fig. 4). Again, the function is denned only for integer values of order. Several such plots of length data made to date seem to yield consistently good fits to a straight line, but the general applicability of the function is not yet established, as in the case of the law of stream numbers. The slope of the regression line ***b*** (Fig. 4) is the exponent in a power function relating the two variables. Marked differences observed in the exponent suggest that it may prove a useful measure of the changing length of channel segments as order changes. Because this is a nonlinear variation, the assumption is implicit that geometrical similarity is not preserved with increasing order of magnitude of drainage basin.

UNIT – 7: DRAINAGE NETWORK: COMPOSITION, LAWS OF STREAM NUMBER AND STREAM LENGTH, NUMBER OF TOPOLOGICALLY DISTINCT CHANNEL NETWORKS, CLASSIFICATION OF LINK TYPES

Development of Rill Systems

The relations of discharge parameters to drainage area were described in connection with the geometry of the drainage basin. Thus far, primary consideration has been given to stable conditions unaffected by evolutionary changes. Such changes will be considered here. Consider initially the smallest channels, ephemeral rills, those channels which carry water only during storms and which most of the time are dry. If we observe the pattern of rills developing on a newly bladed road cut, the first thing that strikes us is the parallel pattern which is initially developed, and the fact that rills bite into the predetermined slope without materially altering the gradient. The head of a rill system may not extend all the way to the watershed divide. A zone near the divide may remain essentially unchanneled. In this zone the depth of overland flow is insufficient to develop an erosive force equal to the forces of cohesion tending to hold the soil particles in place. The processes by which parallel shoestring rills on a fresh surface become integrated into a drainage net are cross-grading and micropiracy-the robbing of a small channel's drainage system by a larger channel. Micropiracy, as explained by Horton (1945, P: 335) is a consequence of the overtopping and breaking down of ridges that initially separate adjacent rills. Two parallel and adjacent rills usually will differ slightly in elevation and depth, the longer being slightly deeper or at a lower elevation or both, compared with the shorter. For where overland flow occurs on a strip of unit width, the depth of flow and thus the ability to erode increases with distance from the drainage divide.

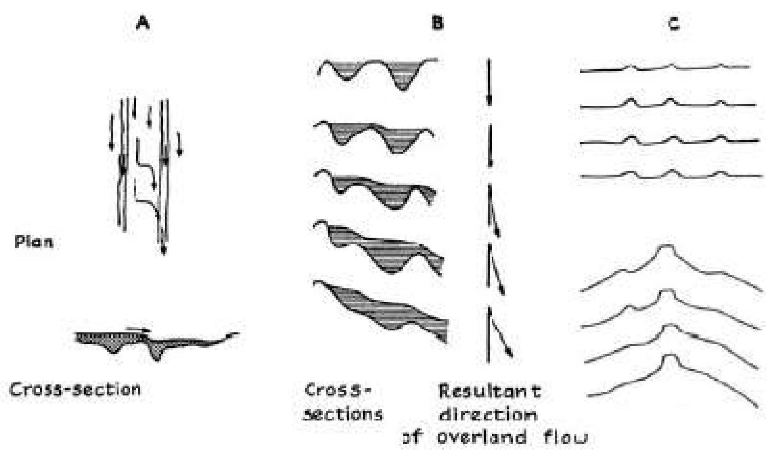


Figure 10-1. *Development of rills on hillslopes and enlargement of rills by cross-grading. [After Horton.]* **A.** *Overland flow over topping divide separating adjacent rills; arrows indicate direction of flow.* **B.** *Successive stages of cross-grading of rills.* **C.** *Contour lines on rilled surface (top) contour lines of the same surface (bottom) after cross-grading.*

During a heavy storm, when overland flow occurs to a depth sufficient not only to fill the adjacent rills but to overtop the intervening divide, the water will tend to develop a lateral component of motion toward the slightly lower rill, and this component across the inter-rill divide will erode it away, as indicated in Fig. 10-1, A. As water is drawn away from one rill by the lower elevation of an adjacent one, the latter will acquire by such micropiracy even more erosive ability at the expense of the former. The interrill divides will progressively be broken down, and a component of flow toward the master or main rill will progressively increase, as indicated in Fig. 10-1, **B.** The development of this component across the main gradient of the general surface was called by Horton "cross grading." As the figure indicates, cross-grading gradually obliterates all the original rill features, and there results a relatively smooth area of overland flow.

The contour lines on the rilled surface and those after cross-grading are suggested in Fig. 10-1, C. The V-shaped re-entrant in the pattern of contours indicates the presence of a component of flow across the main gradient of the hillslope as a result of the cross-grading. It is this new component of slope that allows a new system of cross-grading and the development of a drainage net. The process illustrated in Fig. 10-1 is meant to apply to a rill system rather than a network of creeks or rivers. This qualification is added because the micropiracy or overtopping of inter-rill divides involves processes that can operate quite rapidly and can be observed. Processes operating in large river basins are believed to involve piracy and capture on a larger scale but they are more difficult to observe. The actual overtopping of interstream divides, or cross-grading, occurs but seldom, however, between channels larger than rills, save where rivers debouch from high mountains onto flatter aggradational plains or weaker rocks below.

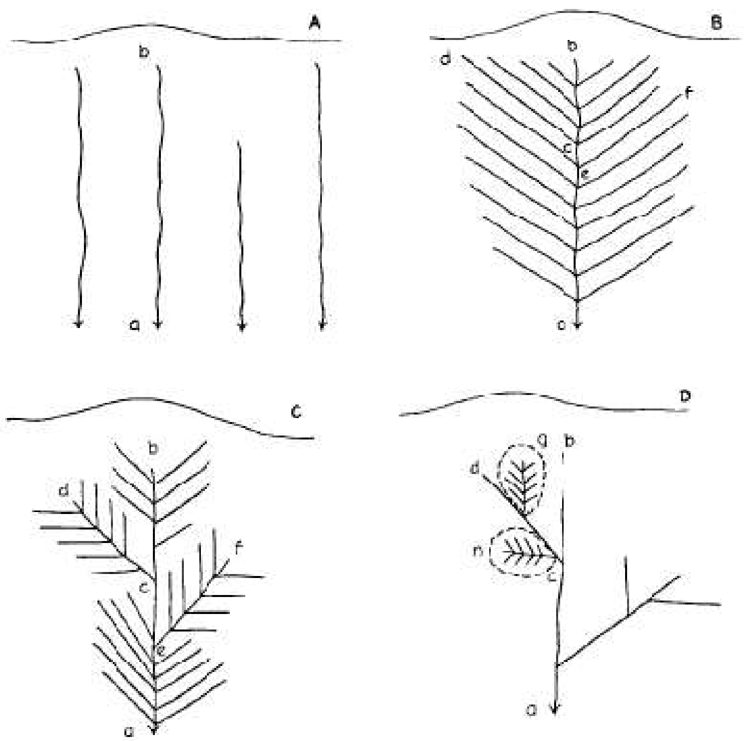


Figure 10-2. *Development of tributary channels and a drainage net by cross-grading. [Partly after Horton.]*

In Fig. 10-2, **A,** a series of new parallel rills are sketched, the longest of which is *abo* By overtopping of divides during overland flow this longest and therefore deepest rill drains water from adjacent rills, which are then obliterated by cross-grading. Figure 10-2, **B,** indicates a single rill *ab,* and the diagonal lines leading toward that rill suggest the direction of overland flow after cross-grading. Along the lines *cd* and *et* the lengths of overland flow are supposedly long enough to attain sufficient depth of overland flow to cause cutting of new rills along those paths. As the rills *cd* and *et* deepen they similarly draw water from adjacent areas and cause cross-grading toward themselves, resulting in directions of overland flow suggested by the sets of parallel lines in Fig. 10-2, **C. In** turn, these new directions of overland flow cause cross-grading toward the positions of the longest length of overland flow and there develops a new pair of rills in positions hand g in Fig. 10-2, **D.** This is the essence of Horton's argument to explain the geometric series by which number and lengths of streams of different orders occur. The sketches in Fig. 10-2 (and those in Horton's discussion of the process) suggest that in this system of rills there are more streams of order $n - 1$ than of order n. Though Horton's brilliant work brought out both the quantitative relations of the geometric series and a general argument in explanation, a somewhat more complete picture of tendencies guiding drainage-net development appears to be provided by the concept of greatest probability, derived from an analogy with entropy. Effect of Longitudinal Profile on Tributary Junctions A key element in drainage-net development and its configuration is that, within limits, the larger of two adjacent stream channels has a tendency to pirate or rob the smaller one of its drainage system. This is relatively easy to visualize in a newly developed system of small rills, owing to the fact that the greater the distance from the watershed divide, the larger will be the amount of water accumulating from a strip of area having unit width. By "a larger amount of water," in this instance, we mean a larger discharge, on the assumption that under a rain of uniform intensity and of sufficient duration each unit of area is contributing the same volume of surface runoff per unit of time. This does not apply to areas even of moderate size, but may hold for areas of a few square yards or even a few acres. In the rill system, the greater the discharge, the greater the depth of flow and rate of rill cutting. At contiguous points in the basins of two adjacent rills, the longer rill will be slightly lower in elevation, and thus water will tend to flow toward that lower elevation when overland flow overtops the inter-rill divide. In larger stream channel systems an analogous relation exists. One of the elements of the Horton system of drainage-net analysis is the relation of channel slope (stream gradient) to stream order. Channel slope decreases exponentially in relation to stream order; channels of larger order have flatter gradients. It would be expected from this generalization that in areas of uniform lithology drainage basins of the same size would be comparable in number and in the lengths of streams of various orders. Further, in similar basins at the same distance from the watershed divide along the principal stream, the channel gradient of the streams would be equal. Even with the variability expected among natural drainage basins, Hack (1957, Figs. 16 and 44) has shown that this holds true. It follows, then, that if in an area of uniform lithology two adjacent stream channels originating at the same elevation differ in length, the longer will lie at a lower elevation. This is shown in Fig. 10-3, where streams *ab* and *cd* are comparable except in length and thus in drainage area. The longitudinal profiles are exactly the same (da' = ba) and originate at the same elevation. But because *cd* is longer than *ab, c* lies at a lower elevation than *a.* If, then, for any reason, the divide between the streams were broken down, water would flow from the smaller toward the larger basin, and the longer or larger stream would tend to capture the drainage area of the smaller.

Probability and the Drainage Network

In the development of a drainage network there are many alternative ways in which rills and channels could develop and unit tributaries meet with their respective master streams. The length of overland flow, the drainage density, the relief, and many other factors that describe the drainage net would depend greatly on the character of the rocks or other materials on which the network is developing, on previous history, and on other parameters. But because there are a very large number of rills, minor channels, and small tributary creeks, it might logically be supposed that in such a large population a mean state might be described. The mean state, or the most probable arrangement within the system, would be that one most likely to be encountered on the average in actual landscapes. Random walks, as we have seen, provide a method by which a model of the most probable state can actually be developed, for it provides a mathematical model which can include the possibility of wide variation among individual examples and one which, on the average, will also yield a picture of the most probable state under the constraints postulated. When precipitation falls on a uniformly sloping plane, an incipient set of rills would develop, oriented generally downhill. The cross-grading that would result and the micropiracy of incipient rills would, as Horton implied" be a matter of chance until the rills deepened sufficiently to become themselves master channels, or master rills. The randomness in the first stages of cross-grading might be approximated by the following model. Consider a set of initial points XI, *X2,* ••• , *X n* (Fig. 10-4) on a line, equidistant from one another at some particular spacing. Let random walks originate at each of these points, and in each unit of time let each random walk proceed from the initial line a unit distance. Let it be specified, however, that each walk may move forward, left, or right at any angle, but may not move backward. The accumulation of moves will produce in time sufficient accumulative departures from the orthogonal to the direction of the original line of points that some paths might meet.

After such a junction only one walk proceeds forward, just as when two stream tributaries join and a single stream proceeds onward. An example of the postulated model is shown in Fig. 10-4, which contains only part of the graphical construction actually made in the trial pictured. The distance any individual random walk will proceed on the average before it meets another can be described by a statistical model called "the gambler's ruin," in which one may compute the most probable number of plays before one player loses all his capital to his adversary. If the capital of each is equal at the beginning of the game and if the size of each wager remains the same, the number of plays expected is one-fourth of the square of the total capital in the game, or *N* ex: *D2,* where *N* is the number of plays and *D* is the total capital.

By analogy, the number of steps of adjacent random walks before one walk is absorbed by or meets another should depend on the square of initial separation distance. But because a walk may deviate either left or right from its general direction, it is less confined than the gambling model, so the meeting may occur as quickly as the first power of the initial separation distance, or *N* ex: *D* (Leopold and Langbein, 1962, p. AI4). A symmetrical model of pairs of streams joining and resultant secondorder streams joining would make the mean spacing of the second-order channels twice that of the first-order, and the mean spacing of third-order four times that of the first-order channels. This geometry is described by an equation, *D* = *d2R* - \ where *D* is mean distance between streams of order *Rand d* is mean distance between first-order streams.

Combining the equations yields the result that the average stream length from one junction to the next varies with the mean interfluve distance in the following form, where *L* is stream length between junctions and *R* is stream order: depending upon whether the length varies as the square or the first power of the distance between streams of a given order. These equations, derived from the mathematical description of the random-walk model, suggest that the logarithms of stream length would vary linearly as functions of stream order and that the mean ratio between lengths of streams differing in order by 1 will be between 2 and 4. This result, derived from considerations of the most probable state, is in agreement with the actual configuration of most channel networks in nature.

Another type of test using random-walk techniques was also developed, using a sheet of rectangular cross-section paper. Each square is assumed to represent a unit area in a developing drainage net, and each square is to be drained. An arrow will be drawn from the center of each square on the graph paper to the center of an adjacent square in a direction specified by pure chance, the direction being decided by a table of random numbers or by some other scheme for producing random choices. Four equal possibilities are postulated for the direction of drainage from each unit squarethe four cardinal directions. There is an equal chance, in other words, that the arrow indicating direction of drainage from a unit square will lead off in any of the four directions. This is subject only to the condition that no internal sinks can be developed.

On such a drainage grid the arrows connect, and a stream network is thus generated, an example of which is shown in Fig. 10-5. Divides are developed and the streams join so as to create drainage areas of various sizes. Such a random-walk pattern may be considered to represent an example of a network in a structurally and lithologically homogeneous area and one in which no differential erosion is accorded to streams of different sizes. The drainage net so developed has statistical properties nearly identical to the properties of natural drainage nets; that is, logarithms of the number of streams of given orders and logarithms of the lengths of given streams are linearly related to stream order itself. This is demonstrated by Fig. 10-6, which refers to the drainage net developed by the random walk pictured in Fig. 10-5. Furthermore, in the random-walk model, stream length is related to the 0.64 power of drainage area, as shown in Fig. 10-6, **B,** a common value for the relationship between length and area in actual streams (Chapter 5). It appears, then, that the logarithmic relations between stream length and numbers of streams with order number is the most probable condition.

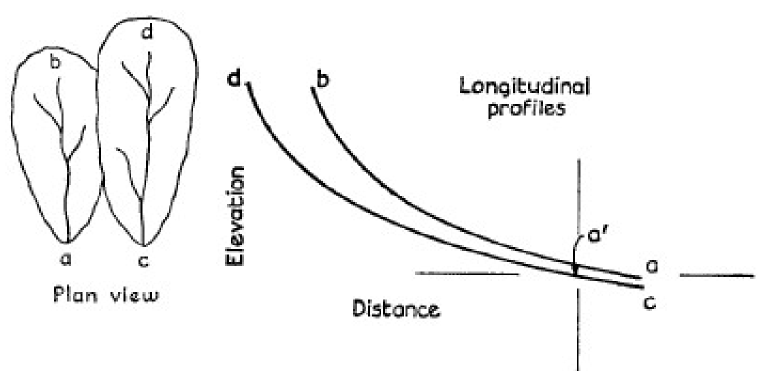


Figure 10-3. *A usual relation of longitudinal profile to stream length in adjacent basins. Profile* da' *is identical to* ba, *rising at the same elevation, but since* dc *is longer,* c *lies at a lower elevation than* a.

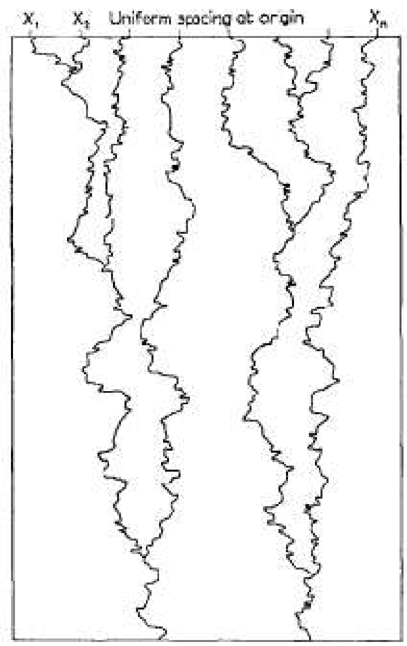


Figure 10-4.

Portion of random walk model of rill or stream network.

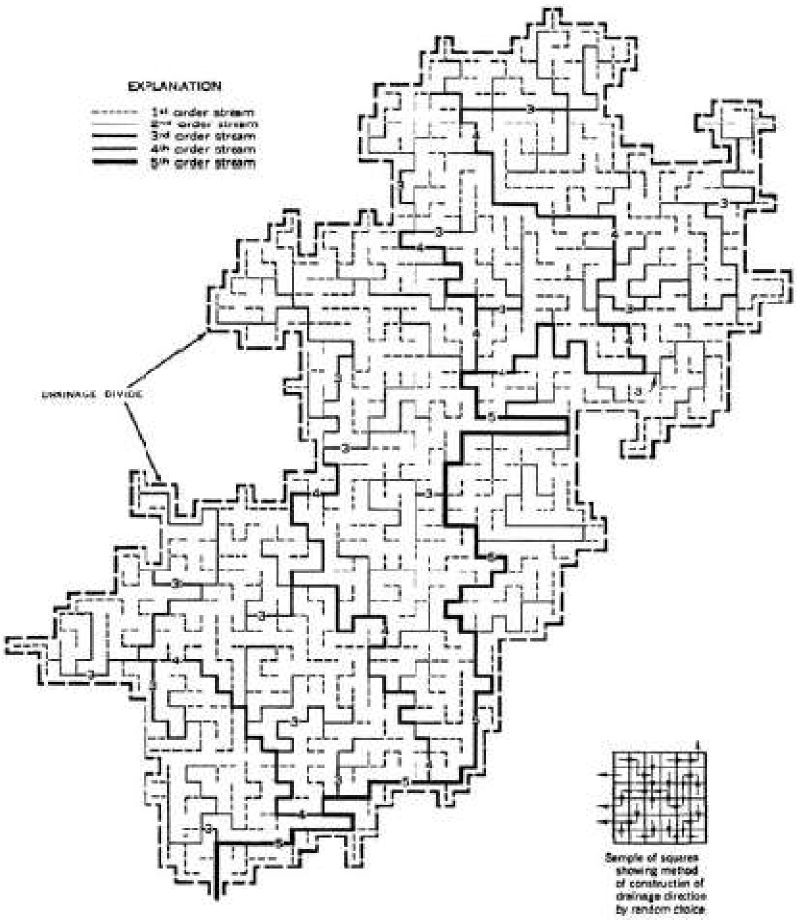


Figure 10-5. *Sample of random-walk drainage network developed on rectangular graph paper.*

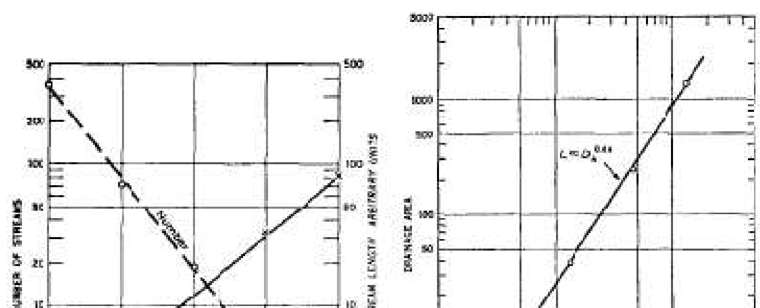


Figure 10-6. *Relations among stream characteristics for drainage net of Fig. 9-5 developed by random walks.* **A.** *Relation of number and average lengths of streams to stream order.* **B.** *Relation of drainage area to stream length.*

Evolution of the Drainage Net

Two different approaches to an initial consideration of the drainage net have been introduced. The rational approach to drainage evolution-suggested by Horton and described here-is based on processes of crossgrading, micropiracy, and effects of overland flow. This indicates that a reasonable drainage net can be built up by applying two principles: (1) when the potential for erosion exceeds the resistance to erosion, a new channel will form or an old one will be extended, and (2) the initial crossgrading between rills is random. Assuming that erosion is a function of angle and length of slope, the general argument about cross-grading leads to the development of a drainage net in which, at least qualitatively, the geometric proportions followed the geometrical properties of natural drainage basins (Horton, 1945, P: 346 et seq.).

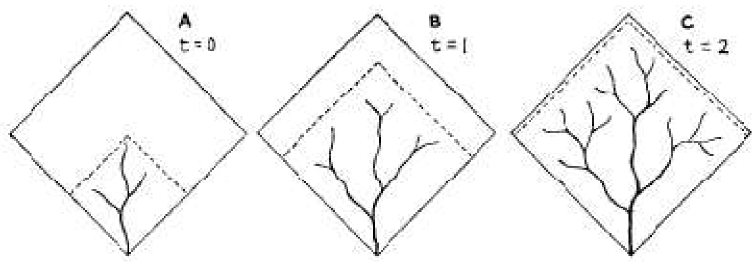


Figure 10-7. *A hypothetical development of drainage net with time; the basin order increases from 2nd order,* **A,** *to 4th order,* **C.**

It is important to recognize that although the geometrical properties of the derived and natural drainage nets are quite similar, the scales are quite different. Even first-order basins in many humid regions are well beyond the incipient cross-grading stage; in all likelihood there are few areas that, on a regional scale, provide a clean slate on which the channel network can develop unhindered or unconstrained by differences in rock type, structure, or initial topographic irregularities. Nevertheless, the density and distribution of stream channels on a large or regional scale does fulfill the physical conditions postulated for the ideal case. The network transports debris derived from the hillslopes without continuously changing its form, and there is a relationship between precipitation, contributing drainage area, and size of first-order stream channels (see Chapter 5).

It has also been shown that these same geometric properties can be "generated" by applying a random process, the random walk, to trace out the paths of lines originating at points along the drainage divide. We have referred to the resulting drainage net as a "most probable configuration." From the relation of channel number to channel length for a large number of drainage basins it can also be concluded that, in general, large basins are geometrically similar to small basins with regard to relations of orders, numbers of streams, and stream lengths. By this we mean that as scale changes both the number of channels and their lengths change proportionately, so that the relation of channel number to channel length shows little scatter. However, the relation of stream length to drainage area changes uniformly but not as the square root of drainage area. The length of a drainage basin increases somewhat faster than the mean width, so basins get relatively longer and narrower with increase in size.

Although both Horton's approach and the random walk give an answer that agrees with observation-that is, the drainage systems so generated have many characteristics in common with real drainage systems-neither tells us whether, with the passage of time, drainage basins do in fact evolve randomly, following the p~stulated principles of erosion. It is also true that the agreement between model and nature is better between statistical or numerical relations than it is between spatial distributions or patterns. To state this in another way, even if one were confident that drainage networks in nature represented the most probable state and the most probable state can be generated by several types of mathematical models, it does not follow that in nature real drainage nets are generated through time in the manner of the model. Although they provide fundamental general principles. with which to begin, neither the rational nor the random walk descriptions indicate how drainage nets develop and change through time.

It is seldom in nature that an area of any size is at any moment unrilled or undissected. **In** nearly all cases there are some inherited characteristics which will influence subsequent development or change and, furthermore, it is rare indeed that lithology and structure are so simple that homogeneity is characteristic of any large area. A river or a drainage basin might best be considered to have a heritage rather than an origin. It is like an organic form, the product of a continuous evolutionary line through time. The most probable state always exists, barring rapid or exceptional changes. This most probable state must always satisfy physical requirements or laws and will therefore be adjusted or altered by the demands of physical in homogeneities. It is useful, then, to inspect examples where either heritage or in homogeneities are partially known or can be minimized.

The geometrical similarity shown by the relation of number and length implies that if a small drainage system begins to develop on a large area *(t* =0 in Fig. 10-7, A), it will ramify in such a way that both channel length and channel number increase (Fig. 10-7, A to C). Considering the total area, *xy,* within the square basin, absolute frequency and drainage density both increase with the passage of time from $t = 0$ to $t = 2$, but if we look only at the area within the dashed lines, we see that each successive basin is geometrically similar to the one preceding it in time.

If we use the concept of stream order, the initial stream as drawn is of order 2, the next of order 3, and the last of order 4. Melton (1958, p. 49) described such a growth model, based on the following logic. Assume that all drainage basins have evolved through time. The innumerable drainage basins whose geometric properties form the statistical basis for the laws of drainage-basin morphology must, in fact, include basins in many stages of development. Yet the statistical data describing these varied drainage basins show no systematic variation which we should expect if different absolute ages were associated with different drainage textures or densities. Melton found (1958, p. 50) that channel frequency for a unit area was, in fact, related not only to channel length but also to basin relief according to the equation where N is the number of channels, L the length of channels, *Ad* the area, and *Re* the total basin relief. Considered as a model of the changes in the drainage net over time where relief remains constant through uplift, the equation implies that the length of the channel system increases simply through an increase in the number of channels. Where relief is reduced, $Re = 1.899/L$; that is, length is inversely proportional to relief. In both cases, however, it is inferred that transformations of drainage nets over time look like existing differences in space.

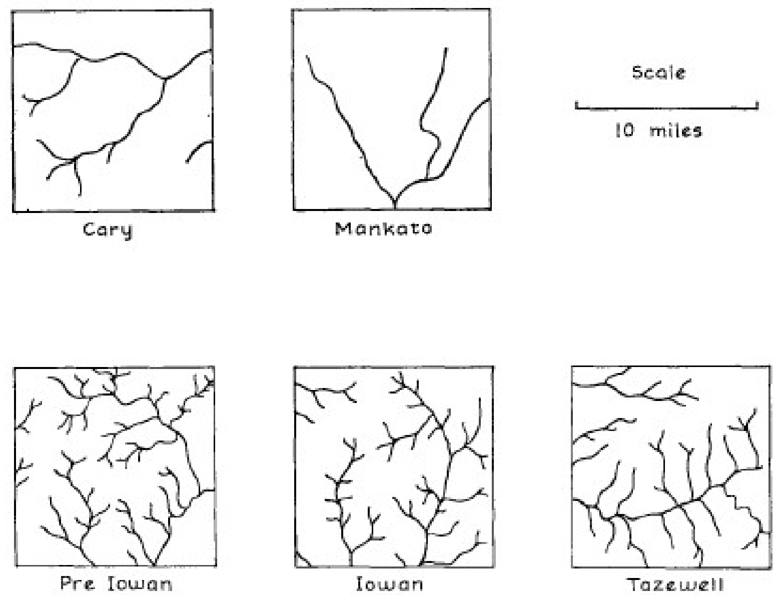


Figure 10-8. *Sample of drainage net on glacial till in Iowa of different age. [After Ruhe, 1952.]*

One of the standard criteria long used by geologists to distinguish glacial tills of different ages is, in fact, the degree of drainage "development." It is postulated that the greater density and the more integrated the drainage pattern, the older the till sheet. The degree of weathering, the superposition of till sheets in stratigraphic sections, and carbon-14 dating correlated with the degree of drainage development have all substantiated this relationship. Thus drainage patterns developed on tills of successive ages can provide an example of progressive changes with increasing time, although variations in initial relief, lithology, and sedimentary composition make cautious comparisons advisable.

Samples of drainage patterns developed on glacial drift of different ages in Iowa are shown in Fig. 10-8 (Ruhe, 1952). For each sample unit area of the five drift sheets, Table 10-1 gives the approximate age in years, the drainage density, and the number of tributaries of successive orders. We recognize that we do not have perfect lithologic identity, and thus to some extent may be substituting space for time; nevertheless, the data do appear to follow the established geologic principle that, expressed numerically, indicates that both drainage density and stream frequency increase with the passage of time. The data are presented in Fig. 10-9, A, as a relation between drainage density and time. The graph implies that the rate of drainage development is most rapid in the first 20,000 years and subsequently levels off to a slower rate. Unfortunately, possible variations in soils and relief, combined with the fact that the dating is not sufficiently precise, do not permit us to state with confidence any more than that channel length and number both increase with time. The leveling off in the rate of increase may reflect the establishment of equilibrium and the filling up of the basin area by stream channels. Figure 10-9, B, from Ruhe's maps, shows that the number of streams per unit area and drainage density (total stream length per unit area) increase systematically with time. In Fig. 10-9, C, drainage density, number of first order tributaries, and total number of tributaries are plotted on semi logarithmic paper as functions of time.

It is interesting to note that the relation between channel number and channel length over time follows the geometric laws which earlier described the relation of number to length for drainage basins of different characteristics in different places. This supports the hypothesis that the spatial distribution is in fact a "growth" model. Unfortunately there are few examples from different environments which illustrate how drainage patterns change with time. Where severe overgrazing or noxious gases from industrial plants have denuded hillslopes of vegetation, accelerated erosion has created new channels and both new and old channels have extended themselves in length.

The network of channels on a portion of the Greenland ice cap show an increase in a 4-year period of about 100% in channel length, and hence in drainage density, accompanied by an increase in the number of channels within the confines of major river valleys and divides on the ice (Fig. 10-10, A and B). As the examples indicate, perpetual transformation of the overall composition of the drainage net does not appear to occur. Development may initially be quite rapid, but the change in pattern after this period is exceedingly slow and the geometrical properties appear then to represent a kind of equilibrium condition appropriate to the available processes, erodibility, and structure of the region. Some aspects of this concept of landform evolution are also considered in Chapter 5. Modes of Drainage Extension I , I

Three modes or physical processes of drainage extension can be cited: (1) surface runoff; (2) subsurface seepage; and (3) headward erosion of scarps or headcuts. Under conditions of surface runoff new first-order channels form where declivity, length of slope, and rainfall intensity create runoff capable of producing shear stresses that exceed the threshold of erosion of the soil. When this happens, the surface cover is broken and rills develop. This process is common on bare roadcuts and on badly managed farmlands. Under such conditions a rill channel may open up anywhere on a slope where local conditions are favorable, and the openings so formed may proceed downslope to already established drainage lines or upslope toward the divide.

Less obvious processes of incipient channel extension have recently been described by Bunting (1961) on moorlands in England. Detailed mapping of the subsurface soil mantle reveals seepage lines containing fine eluviated material and relatively deep layers of humus. These seepage lines have no topographic expression and are distinguishable at the surface only by their moist appearance at certain times of the year. The presence of moisture, and its slow downward movement evident in the soil profile, promotes continuous corrosion of the bedrock (Fig. 10-11, A). In some areas the seepage lines connect directly with the perennial first order drainage channels. Thus seepage promotes both a lowering of the surface of the bedrock by corrosion and an extension of the drainage network. Because water availability is limited at the crests of divides, continuous lowering of the adjacent hillslopes may leave isolated remnants of rock (tors) perched upon the drainage divides.

A third process of drainage extension-the headward retreat of vertical or nearly vertical channel scarps-has been described in many regions. A vertical face or drop in the bed of a stream channel, often called a headcut, may be created in a variety of ways. Faulting, lowering of the mainstream at the junction of a tributary, local concentrations of inflow, or topographic and stratigraphic irregularities may cause incipient gullies and headwalls to form. Once formed, a headcut may migrate upvalley, thus extending the drainage net within the drainage area. The mechanics of the process of headward migration of scarps is considered more fully in connection with channel degradation (Chapter 11). Gullies may extend themselves headward by basal sapping, aided on occasion by moisture or weaker strata at the base of the scarp, by concentrated erosion by surface flow in a plunge pool, or by combinations of these mechanisms as yet poorly understood.

Not infrequently gully formation is an aspect of accelerated erosion produced by changes in land use or climate. Such changes may produce increased runoff, greater seepage, increased erodibility through reduced binding of the soil, or many other changes and combinations of effects. Headward migration of channel scarps may increase the drainage density. In many cases the drainage net might consist of swales without marked channels, prior to trenching by gullies. Such swales are not unlike the subsurface seepage lines mentioned above, and the difference in composition of the drainage net before and after migration of the headcuts is in part a function of the way in which lines or channels in the network are defined. Where gullies form on slopes, however, these may constitute more obvious new channels, although the locus of formation may have been a modest depression or swale. In some regions the principal means of increasing the network of drainage channels may be through headward migration of gullies.

The rate of retreat of a gully headcut appears to be related to both runoff and moisture content of the soil. Repeated cycles in the southwestern United States show clearly that this process may create ramified drainage patterns comparable to those apparently created by other processes elsewhere. Descriptions of erosional processes on slopes reveal that although specific processes may be dominant in some climatic and physiographic environments, a variety of processes are usually operative in every region. One should expect, therefore, in most drainage basins, that surface runoff, seepage, and headcut retreat all participate in developing the drainage system.

Where moisture is deficient, seepage may occur only after rain, as it does in the free face of headcuts in semiarid regions. Where landslides are frequent, they may intermittently obliterate evidence of stream channels until surface runoff is again sufficient to trench the surface of the debris provided by the slides. On a very much smaller scale, for example, Schumm (1956) observed that rills formed on the surface of a fill at Perth Amboy, New Jersey, during the spring, summer, and fall, and were obliterated each winter by frost action.

Limits of Drainage Development

At present there is insufficient evidence to tie drainage evolution over long periods of time to the detailed operation of specific processes. However, the operation of several basic mechanisms is apparent in the landscape. Where the surface of land in pasture, for example, is burned, trampled, or otherwise destroyed, resistance to erosion is lowered, rills form by surface runoff, and these in turn grow into gullies. Intrenchment of gullies may produce steep walls. At points of weakness, where seepage moisture is concentrated, portions of the wall may begin to retreat headward. These processes have been observed in virtually every environment on the globe. All contribute to the development of the drainage system. Drainage development, of course, cannot be divorced from hillslope development. Assuming a constant drainage area, it is clear that as dissection proceeds and channel numbers and lengths increase, and if the crests of divides are not lowered, slopes will increase in steepness. As divides are lowered, slopes may decline in steepness, although changes in the declivity of slopes over very long periods of time may be exceedingly slow and will vary depending upon the relative rate of crest and channel lowering.

One may well ask, is there a limit to drainage development? If so, what is it like? The answer lies in the way one looks at or defines the dynamic system and also in the time scale considered. An open system in steady state may, in one sense, be constant in form, though energy is constantly put into the system and degraded to heat within it. Such a view is both theoretic and limiting. The utility of the concept of the open system in steady state is that in nature close approximations to the steady state do occur. At the same time, however, landforms do evolve or change gradually through time while maintaining a steady-state form. It was pointed out in connection with the concept of entropy or the most probable distribution or form that the steady state is closely approximated by the average of a number of examples in a given population. For example, in a number of similar drainage basins in the same climate and on the same materials, the average characteristics of the group may describe very closely the steady state of an open system. With random walks, furthermore, only a relatively small number of trials were necessary to obtain average results that closely approximated the average of a large number of trials. This finding suggests, even though it does not prove, that random processes operating within physical and hydraulic constraints develop rather quickly the same characteristics that obtain after a much longer period of time. Thus there is plenty of play or room within the concept of the steady state of an open system for the gradual evolution and slow change of average conditions through time, maintaining all the while a close approximation to the condition of dynamic equilibrium with conditions existing at any point in time. Given topographic variability or relief and erosional processes operating over geologic time, uplift and denudation are not in perfect equilibrium. It should be expected that relief will change and in association with it hillslope and channel form. These changes, however, may be exceedingly slow and much overshadowed at any point in time by differences attributable to variations in processes or structure. Thus it is not only possible but profitable in the analysis of the origin and behaviour of landforms to consider landscape as a system representing a steady state or dynamic equilibrium.

Drainage patterns and hillslopes show a high degree of adjustment of form and pattern to the tasks of handling variable quality and quantities of debris under diverse environmental conditions. This suggests simply that in the long run, adjustment of form to process in the landscape must be relatively rapid, and further, that the rate of change of many forms, including the drainage net, is slow save under unusual circumstances, such as man creates. As illustrated by the studies of Hack and Goodlett (1960) in the Shenandoah Valley, catastrophic rainfall and floods can create new stream channels with each catastrophe. The number of stream channels does not appear to be ever increasing. Rather, some channels must be erased or filled as others develop, and the resultant drainage net continues to provide channels sufficient in frequency and size to carry off the hil1slope debris and runoff.

The adjustment of form and pattern to structure and process is readily attained and changes slowly. Minor deviations from this adjusted or quasi equilibrium condition are difficult to recognize. As statistical studies show, many parameters used to describe the profile and drainage pattern are relatively insensitive to departures from the average or most probable form.

UNIT - 8: ANALYSIS OF DRAINAGE BASIN: BASIN SHAPE, FORM FACTOR (HORTON, 1932), ELIPTICITY INDEX (STODDART, 1965), CIRCULARITY INDEX (MILLER, 1953), INFLUENCE OF BASIN SHAPE ON HYDROLOGICAL REGIME; BASIN SIZE: AREA RATIO AND LAW OF BASIN AREA, LAW OF ALLOMETRIC GROWTH AND BASIN AREA; RELATION OF BASIN AREA WITH STREAM NUMBER AND LENGTH

Drainage basin areas—Area of a given w a t e r s h ed or drainage basin, a p r o p e r t y of the square of length, is a p r i m e d e t e r m i n a n t of total runoff or sediment yield and is normally eliminated as a v a r i a b l e by reduction to u n i t area, as in annual sediment loss in acre-feet per s q u a r e mile. I n order t o compare drainage basin areas in a meaningful way, it is necessary t o compare basins of t h e same order of m a g n i t u d e . T h u s , if we measure t h e areas of drainage basins of the second order, we are measuring corresponding elements of t h e systems. If approximate geometrical similarity exists, the a r e a measurements will then be indicators of the size of t h e l a n d f o rm u n i t s , because areas of similar forms are r e l a t e d as t h e square of t h e scale r a t i o.

Basin area increases exponentially w i t h s t r e am order, as s t a t e d in a l aw of a r e a s ***[Schumm,*** 1956, p .606], p a r a p h r a s i n g H o r t o n ' s l aw of s t r e am lengths. Schumm [1956, p . 607] h a s shown histograms of t h e areas of basins of the first and second orders a n d of p a t c h e s of ground surface too small t o h a ve channels of t h e i r own. Basin a r e a d i s t r i b u t i o n s a re strongly skewed, b u t this is largely corrected by use of log of a r e a . Area is measured by p l a n i m e t er from a topographic map, hence represents projected, r a t h e r t h a n t r u e surface area. E s t i m a t i on of true surface area has been a t t e m p t e d where surface slope is known ***[Strahler,*** 1956a, p . 579].

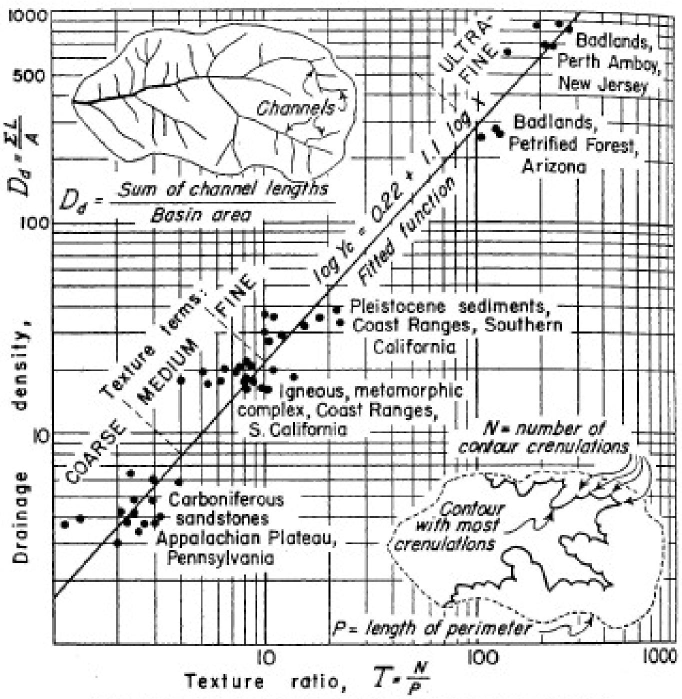


FIG. 5 – Definitions of drainage density and texture ratio (Strahler, 1954a, p. 348)

Drainage density and texture ratio—An important indicator of t h e linear scale of landform elements in a drainage b a s i n is drainage density, denned by ***H** or ton* [1945, p . 283]. T h e u p p e r left-hand corner of F i g u r e 5 shows t h e definition of drainage d e n s i ty as the sum of the channel lengths divided by b a s i n area. D i v i s i o n of length b y a r e a thus yields a number with t h e dimension of inverse of length. I n general, then, as t h e drainage density number increases, the size of individual drainage units, such as the first-order drainage basin, decreases p r o p o r t i o n a t e l y .
F i g u r e 5 shows the relation between drainage d e n s i t y and a related index, the texture ratio, defined by *Smith* [1950]. Because the contour inflections on a good t o p o g r a p h i c map indicate the existence of channels too small to be shown by s t r e am symbols, their frequency is a measure of closeness of channel spacing and hence also correl a t e s w i t h d r a i n a g e d e n s i t y. D r a i n a g e density is scaled logarithmically on t h e o r d i n a t e of F i g u r e 5. T h e grouped points in the lower left-hand corner of the graph represent basins in r e s i s t a n t , massive sandstones. Here the streams a r e widely spaced a n d density is low. The next g r o u p of p o i n t s e n c o u n t e r e d represents typical densities in deeply weathered igneous and metamorphic rocks of the California coast ranges. In t h e extreme upper right are points for badlands, where d r a i n a g e d e n s i t y is from 200 to 900 miles of channels per square mile ***[Smith,*** 1953; ***Schumm,*** 1956, p- 612].

Because of its wide r a t i o of v a r i a t i o n , drainage density is a number of p r i m a r y importance in landform scale analysis. One might expect that sediment yield would show a close p o s i t i v e relationship with drainage density. A r a t i o n a l theory of the relation of d r a i n a g e d e n s i t y t o erosion intensity, predicting the morphological changes to be expected when ground surface resistance is lowered by land use, h a s b e e n outlined b y ***Strahler*** [19565],

Constant of channel maintenance—Schumm [1956, p. 607] has used the inverse of drainage density as a p r o p e r t y termed c o n s t a n t of channel maintenance. I n F i g u r e 6 t h e l o g a r i t hm of basin area (ordinate) is t r e a t e d as a function of l o g a r i t hm of total s t r e am channel length (abscissa). Stream length is cumulative for a given order a n d includes all lesser orders; i t is t h u s t h e t o t a l channel length in a watershed of given order. L e n g t h in this case is projected to the horizontal p l a n e of the map; true lengths would be obtained by applying a correction for slope. An individual p l o t t e d point on t h e g r a p h represents a given s t r e am order in the watershed, as numbered 1 t h r o u g h 5. Using d a t a of the three examples given b y Schumm, t h e sets of points fall dose to a s t r a i g h t line of 45° slope; t h u s t h e relationship is t r e a t e d as linear even t h o u g h plotted here on log-log p a p e r . If the l o g a r i t hm of t h e intercept is read a t log s t r e am l e n g t h = 0, a n d t h e a n t i - log of this intercept is t a k e n , we o b t a i n the con- t a n t of channel maintenance C which is a c t u a l ly t h e slope of a linear regression of area on length. The value of **C** - 8.7 in t h e P e r t h Amboy b a d lands means t h a t on t h e average 8.7 sq ft of surface a r e required to maintain each foot of channel length. I n t h e second example, Chileno Canyon in t h e California San Gabriel M o u n t a i n s , 316 sq ft of surface a r e r e q u i r e d to m a i n t a i n one foot of channel length.

The constant of channel maintenance, w i t h the dimensions of length, is thus a useful means of indicating t h e relative size of landform u n i t s in a drainage basin and has, moreover, a specific genetic connotation.

Maximum valley side slopes—Leaving now the drainage network and what might be classified as planimetric or areal aspects of drainage basins, we t u r n to slope of the ground surface. This brings i n t o consideration the aspect of relief in d r a i n a ge basin geometry. One significant indicator of the over-all steepness of slopes in a watershed is t he maximum valley-side slope, measured a t i n t e r v a ls along the valley walls on t h e steepest p a r t s of t he contour orthogonals running from divides to adjacent s t r e am channels. Maximum valley-side slope has been sampled by several investigators in a wide v a r i e t y of geological and climatic environments ***[Strahler,*** 1950; ***Smith,*** 1953; ***Miller,*** 1953; ***Schumm,*** 1956; ***Coates,*** 1956; ***Melton,*** 1957]. W i t h i n - a r e a v a r i a n ce is relatively small compared with between-area differences. This slope statistic would therefore seem to be a valuable one which might r e l a te closely to sediment production.

Mean slope curve—Another means of assessing t h e slope p r o p e r t i e s of a drainage basin is t h r o u gh t h e mean slope curve ***[Strahler,*** 1952, p. 1125- 1128]. This requires the use of a good contour topographic map. T h e problem is t o e s t i m a t e the average, or mean slope of t h e b e l t of ground surface lying between successive contours. This may be done by measuring the area of each contour belt with a planimeter and dividing this area b y the length of t h e contour belt to yield a mean width.

The mean slope will then be t h a t angle whose tangent is the contour interval divided by the mean belt width. Mean slope of each contour interval is p l o t t e d from summit point to basin mouth. Curves of this type will differ from region t o region, depending upon geologic s t r u c t u r e and t h e stage of development of t h e drainage system. If the mean slope for each contour belt is weighted for per cent of t o t a l basin surface area, i t is possible t o a r r i v e a t a m e a n slope v a l u e for t h e surface of t h e w a t e r s h e d a s a whole.

Slope maps—Another means of determining slope conditions over a n e n t i r e g r o u n d surface of a w a t e r s h e d is t h r o u g h the slope map *[Strahler,*1956a]. (1) A good topographic m a p is t a k e n . (2) On this m a p t h e slope of a short segment of line normal to t h e t r e n d of t h e c o n t o u r s is d e t e r m i n ed a t a large n u m b e r of p o i n t s . T h e s e m a y b e recorded as t a n g e n t s or sines, depending upon t h e k i n d of m a p desired. (3) These readings are contoured w i t h lines of equal slope, h e r e called i s o t a n g e n t s. (4) T h e areas between successive iso t a n g e n t s a re measured with a p l a n i m e t e r a n d t h e a r e a s summed for each slope class. (5) T h i s yields a slope frequ e n c y percentage distribution. Because the e n t i r e ground surface h a s b e e n analyzed, t h e m e a n , s t a n d a r d deviation, and v a r i a n c e are t r e a t e d as p o p u l a t i o n p a r a m e t e r s , a t least for purposes of comparison w i t h small samples taken a t r a n d om f r om t h e same area. Lines of equal sine of slope, or isosines, m a y also b e d r a w n . T h e i n t e r v a l between isosines on t h e m a p becomes t h e s t a t i s t i c a l class on t h e h i s t o gram. Sine values are designated as *g* values because t h e s i n e of slope r e p r e s e n t s t h a t p r o p o r t i on of t h e acceleration of g r a v i t y acting in a downslope direction parallel w i t h t h e g r o u n d surface.

Rapid slope sampling—The construction of slope maps a n d their areal measurement is ext r e m e l y time-consuming. E x p e r i m e n t s h a v e shown t h a t essentially the same information can be achieved by r a n d om point sampling *[Strahler,* 1956a, p. 589-595]. Both random coordinate sampling and g r i d sampling h a v e been tried. In t h e r a n d o m - c o o r d i n a t e m e t h o d a sample s q u a r e is scaled in 100 l e n g t h u n i t s p e r side. F r o m a t a b l e of r a n d om n u m b e r s t h e coordinates of sample p o i n ts a r e drawn for w h a t e v e r sample size is desired. T h e g r id m e t h o d does much t h e same thing, b u t is n o t flexible as to s a m p l e size. P o i n t samples, which are easy to take, were compared with the frequency d i s t r i b u t i o n measu r e d from a slope map. N o t e w o r t h y is t h e ext r e m e l y close agreement in means and v a r i a n c e s, a n d even in t h e f o rm of t h e f r e q u e n c y d i s t r i b u t i o n s, including a m a r k e d skewness. Tests of sample v a r i a n c e and mean are discussed by *Strahler* [1956a]. *Chapman* [1952] has developed a method of analyzing b o t h azimuth and angle of slope from contour topographic maps. Although based on p e t r o f a b r i c m e t h o d s and designed largely for u se in geological analysis of t e r r a i n , t h e method mjght be applied t o a w a t e r s h e d as a means of assessing b o t h slope steepness a n d orientation simultaneously.

Relief ratio—*Schumm* [1956, p . 612] h a s devised a n d applied a simple s t a t i s t i c , the relief ratio denned as t h e r a t i o b e t w e e n t o t a l basin relief (that is, difference in elevation of basin mouth and summit) and basin length, measured as the longest dimension of the drainage basin. In a general way the relief ratio indicates overall slope of the watershed surface. It is a dimensionless number, readily correlated with other measures that do not depend on total drainage basin dimensions. Relief ratio is simple to compute and can often be obtained where detailed information on topography is lacking.

Schumm [1954] has plotted mean annual sediment loss in acre feet per square mile as a function of the relief ratio for a variety of small drainage basins in the Colorado Plateau province [Fig. **7].** The significant regression with small scatter suggests that relief ratio may prove useful in estimating sediment yield if the parameters for a given climatic province are once established.

Hypsometric analysis—Hypsometric analysis, or the relation of horizontal cross-sectional drainage basin area to elevation, was developed in its modern dimensionless form by *Langbein* and others [1947]. Whereas he applied it to rather large watersheds, it has since been applied to small drainage basins of low order to determine how the mass is distributed within a basin from base to top ***[Strahler*** 1952; ***Miller,*** 1953; ***Schumm,*** 1956; ***Coates,*** 1956]. Figure 8 illustrates the definition of the two dimensionless variables involved. Taking the drainage basin to be bounded by vertical sides and a horizontal base plane passing through the mouth, the relative height is the ratio of height of a given contour ***h*** to total basin height ***E.*** Relative area is the ratio of horizontal cross-sectional area ***a*** to entire basin area ***A.*** The percentage hypsometric curve is a plot of the continuous function relating relative height ***y*** to relative area ***x.*** As the lower right-hand diagram of Figure 8 shows, the shape of the hypsometric curve varies in early geologic stages of development of the drainage basin, but once having attained an equilibrium, or mature stage (middle curve on graph), tends to vary little thereafter. Several dimensionless attributes of the hypsometric curve are measurable and can be used for comparative purposes. These include the integral, or relative area lying below the curve, the slope of the curve at its inflection point, and the degree of sinuosity of the curve. Many hypsometric curves seem to be closely fitted by the model function shown in the lower left corner of Figure 8, although no rational or mechanical basis is known for the function. Now that the hypsometric curves have been plotted for hundreds of small basins in a wide variety of regions and conditions, it is possible to observe the extent to which variation occurs. Generally the curve properties tend to be stable in homogeneous rock masses and to adhere generally to the same curve family for a given geologic and climatic combination.

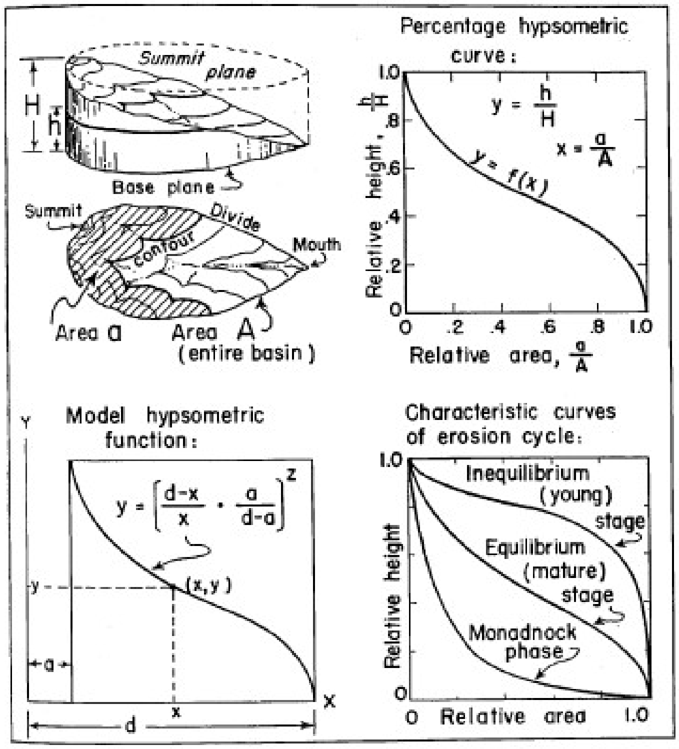


Fig. 8 – Method of hypsometric analysis (Strahler, 1954a. p. 353)

Conclusion-This chapter has reviewed briefly a variety of geometrical properties, some of length dimension or its products, others dimensionless, which may be applied to the systematic description of drainage basins developed by normal processes of water erosion. Among the morphological aspects not mentioned are stream profiles and the geometry of stream channels. These, too, are subject to orderly treatment along the lines suggested. The examples of quantitative methods presented above are· intended to show that, complex as a landscape may be, it is amenable to quantitative statement if systematically broken down into component form elements. Just which of these measurements or indices will prove most useful in explaining variance in hydrological properties of a watershed and in the rates of erosion and sediment production remains to be seen when they are introduced into multivariate analysis. Already there are definite indications of the usefulness of certain of the measures and it is only a matter of continuing the development of analytical methods until the most important geomorphic variables are isolated.

1.7 SELF-ASSESSMENT TEST

Discuss about various components of fluvial system.

Discuss about different variable of fluvial system. What are the different scales in fluvial morphometry?

Discuss about various aspects of drainage evolution.

1.8 SUMMARIES AND KEY POINTS

The following topics have been be discussed in the present section -

- Scope and significance of fluvial geomorphology;
- Fluvial system: concepts, components, input output with stores of material and energy;
- Variables of fluvial system: internal and external, adjustable and controlling; Scales in fluvial geomorphology: times and space scales;
- Channel initiation: theory of overland flow, theory of sub-surface flow;
- Flow regime: hydrological pathways, measurement of stream flow and annual flow regimes; Classification of natural streams by D.L. Rosgen
- Drainage evolution: parallel development and coalescence, headward extension and branching, lateral expansion;
- Drainage network: composition, laws of stream number and stream length, number of Topologically Distinct Channel Networks, classification of link types;
- Analysis of drainage basin: Basin shape, form factor (Horton, 1932), elipticity index (Stoddart, 1965), circularity index (Miller, 1953), influence of basin shape on hydrological regime; Basin size: area ratio and law of basin area, law of allometric growth and basin area; Relation of basin area with stream number and length

1.9 STUDY TIPS

Fundaments of fluvial morphometry – R. Charlton

Fluvial Processes in Geomorphology – Leopold, Wolman and Miller

GROUP – B: MECHANISM OF FLUVIAL PROCESSES [CREDIT: 2]
(MARKS - 50: INTERNAL EVALUATION – 10, SEMESTER-END EXAMINATION - 40)

2.1 INTRODUCTION

The present section discusses mechanics of flow, velocity and resistance, mechanics of fluvial erosion, river transportation, sediment load, mechanics of sediment deposit and sediment deposition.

2.2 LEARNING OBJECTIVES

- Mechanics of flow: uniform and non-uniform, steady and unsteady, laminar and turbulent, tranquil and rapid;
- Velocity and resistance: factors affecting velocity, flow resistance and viscosity;
- Stream power and energy;
- Mechanics of fluvial erosion: threshold of erosion, entrainment and bed erosion, headward erosion and channel lengthening, bank erosion by fluvial processes;
- River transportation: process of particle entrainment, initiation of sediment transport, selective theories of sediment transport, theories of equal mobility transport, transport of bed load and suspended load;
- Sediment load: nature and types of sediment load, sources of sediment load;
- Mechanics of sediment deposit: Channel competence, capacity and efficiency;
- Sediment deposition - processes, nature and characteristics of fluvial deposits; Flood plain and deltaic plain deposits

2.3 ASSESSMENT OF PRIOR KNOWLEDGE

Fundamental concepts of geomorphology will be discussed

2.4 LEARNING ACTIVITIES

Identification and demarcation of drainage basin will be done. Sediment analysis may be done.

2.5 FEEDBACK OF LEARNING ACTIVITIES

Discussion on various aspects of fluvial morphometry will be helpful

2.6. EXAMPLES AND ILLUSTRATIONS:

UNIT – 9: MECHANICS OF FLOW: UNIFORM AND NON-UNIFORM, STEADY AND UNSTEADY, LAMINAR AND TURBULENT, TRANQUIL AND RAPID

FLOW BEHAVIOUR

Subcritical, critical and supercritical flow

The unsteady, gradually varied flow in most natural channels is **subcritical**. However, another type of flow behaviour, **supercritical flow**, is also observed. Within supercritical flows, turbulent mixing is less intense, with less deviation from the main downstream direction of flow. As a result, supercritical flows move rapidly and efficiently through the channel. They may overshoot tight bends and can also be highly erosive (Kay, 1998). The different types of flow behaviour can be predicted by calculating the ratio between the inertial and gravitational forces. The inertial force (see p. 82) is given by *v2/d*; where *v* is the flow velocity, and *d* is its depth. The gravitational force is the acceleration due to gravity, *g*. The ratio between these forces is usually expressed in the form:

$$Fr = \frac{v}{\sqrt{gd}}$$

where Fr □□Froude number, *v* □□velocity, *g* is gravitational constant and *d* is depth. At Froude numbers less than 1 the gravitational forces dominate and the flow is subcritical. Conversely, when the inertial forces dominate, at Froude numbers greater than 1, the flow is supercritical. In rare cases, where the Froude number is equal to 1, the flow is described as being **critical**, or transitional. Figure 6.6 shows what happens when there are transitions between these two types of flow. A **hydraulic drop** occurs when subcritical flow changes to supercritical flow.

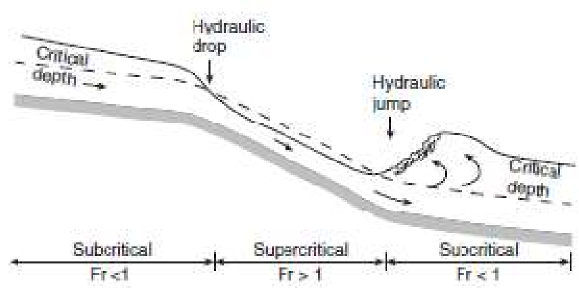


Figure 6.6 Illustration of hydraulic drop and hydraulic jump.

In this example, the increase in channel slope increases the flow velocity, resulting in a reduction in depth (the hydraulic drop). This can be seen in Plate 6.1, as water flows smoothly and rapidly down the steep slope of a weir. At the base of the weir, the flow changes back to subcritical, forming a

hydraulic jump. A breaking wave indicates where this transition occurs. The sudden change in flow conditions at the hydraulic jump is caused by the decrease in slope at the base of the weir. Associated with this is a decrease in velocity and an increase in depth. The high velocity flow has considerable inertia and continues along the bed of the river before it is 'pulled' up to the surface and into the breaking wave. Turbulence is increased in this zone because of shear between the downstream and upstream movements of water.

Flow separation

Flow separation occurs where there are irregularities in the boundary. Examples include abrupt changes in bank orientation, sharp bends and obstructions at the bed such as large boulders. This results in the detachment of the boundary layer, which continues on in the direction of the flow but as a free shear layer. Figure 6.7(a) illustrates flow separation caused by a boulder on the channel bed. In the flow separation zone, between the free layer and the boundary, is a 'bubble' of slow moving, recirculating fluid. The large difference in velocity between the fast moving shear layer and slowly recirculating flow in the separation zone means that large shear stresses develop. The resultant transfer of momentum leads to the free layer becoming unstable a certain distance downstream from the separation point, where it reattaches to the boundary. Increased turbulence is created by flow separation and results in a **wake** downstream from the object. Figure 6.7 (b and c) shows the effect of flow separation around an obstruction and at a bend. Since flow separation affects shear stress distributions, it also influences processes of sediment erosion and deposition.

Laminar and turbulent flow

When considering the internal structure of fluid flow, a distinction is made between two quite different types of flow: laminar and turbulent. The British engineer Osborne Reynolds first demonstrated the existence of these two types of flow in his well known experiments on flows through pipes, carried out in the 1870s and 1880s. By injecting a thin stream of coloured dye into the water, Reynolds was able to observe patterns of movement within the flow. At low flow velocities, the dye was seen to travel as a single thread in a straight line through the tube, and was described by Reynolds as *direct* flow (now known as **laminar** or **viscous flow**). In laminar flows, the fluid moves as a series of layers, which slide over one another. This can be visualised as being somewhat similar to the way in which a pack of cards slide over each other when a shear stress is applied. Highly viscous fluids, such as oil or treacle, tend to exhibit laminar flow because of their high resistance to deformation. This can be seen from the 'smooth' way in which these fluids flow over a surface when gently poured. Water has a relatively low viscosity, so laminar flow only occurs at very low flow velocities.

Reynolds found that a second, very different, type of flow occurred at higher velocities. In contrast to laminar flows, a series of horizontal and vertical swirling motions developed, dispersing the dye throughout the flow. Described as *sinuous* by Reynolds, this flow behaviour was subsequently termed **turbulent flow** by Lord Kelvin. Within the three-dimensional body of flow, movement can be in any direction: vertically up or down, sideways, upstream, downstream, or any combination of these. Reynolds found that as flows changed from laminar to turbulent, a transitional flow-type developed, with the turbulence intensity increasing as the flow became fully turbulent. From these experiments it was clear that two different types of flow behaviour existed. What was not so clear was how the transition between these flows could be predicted, as velocity is only one of a number of variables that control flow behaviour. Reynolds conducted further experiments using different fluids, and pipes with varying diameters. From these experiments, he derived an equation to define the transition from laminar to turbulent flow as a function of a single parameter, the **Reynolds number (Re)**. At low Reynolds numbers laminar flow occurs and at high values, turbulent flow. The concept of a Reynolds number is fundamental to much of modern fluid dynamics. It is calculated using the **Reynolds equation**, which expresses the ratio between **inertial** and **viscous forces** acting on the fluid. The inertia of an object – in this case a body of flowing water – is defined by its mass. Inertia determines how difficult it is to set something in motion (here: to initiate flow) but also how difficult it is to stop it, slow it down or change its direction once it has started moving. The greater the mass of water, the more inertia it has. Fluids that are denser than water (e.g. mercury) have more inertia because their mass per unit volume is greater. The inertial forces also increase with velocity.

Acting against the inertial forces are viscous forces, which resist fluid deformation and flow. The viscosity of a fluid is determined by its structure at the molecular level, as work must be done to move the molecules past one another. Factors such as the regularity of molecular shapes and the strength of attraction between molecules affect the way in which the fluid responds to deformation. The more viscous a fluid is, the more it resists deformation and the less easily it flows.4

The Reynolds number, Re, is given by:

where: Re □□Reynolds number, *v* □□mean flow velocity,

R □□hydraulic radius, r □□fluid density and n (Greek letter nu, which confusingly looks rather like a ‘*v*’) □ kinematic viscosity. At low Re numbers (less than 500), the viscous forces dominate and flow is laminar. Where the inertial forces are dominant (at Re numbers greater than 2,100), the inertia of the flowing water is much more significant than the viscous forces resisting that movement and turbulent flow occurs. The transition between laminar and turbulent flow occurs between Re values of 500 and 2,000. The Reynolds number is dimensionless: it does not have units. When velocity (in m s−1) and hydraulic radius (m) are multiplied together, the resultant units are m2 s−1. The units for kinematic viscosity are also m2 s−1, so when the Reynolds number is calculated the units cancel out.

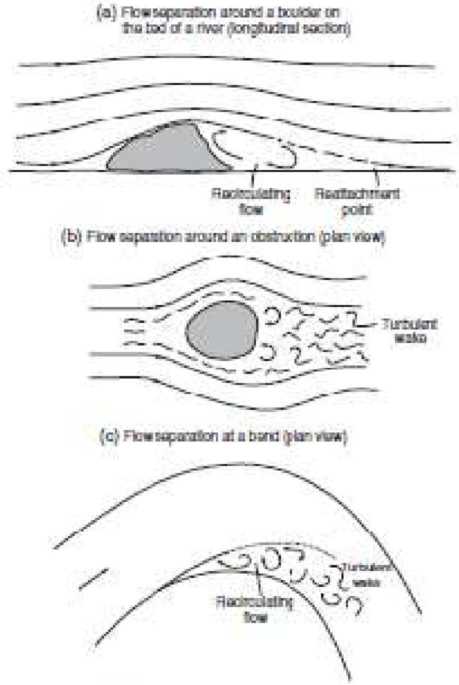


Figure 6.7 (a) Flow separation around a boulder on the bed of a river. (b) Flow separation around an obstruction. (c) Flow separation at a bend. (b) and (c) adapted from Morisawa (1985).

UNIT -10: VELOCITY AND RESISTANCE: FACTORS AFFECTING VELOCITY, FLOW RESISTANCE AND VISCOSITY

INTRODUCTION TO FLOW IN RIVER CHANNELS

Forces driving and resisting the flow of water

A force is anything that moves an object, or causes the speed or direction of a moving object to change. Forces are vector quantities, which means that they have both magnitude (size) and direction. The unit of force is the newton (N), and force magnitude is defined by the mass of the object and the acceleration produced.1 Forces are always mutual. In other words, if a force is exerted on an object, the object will react with an equal and opposite force. In most situations, several forces are involved, so the balance between driving and resisting forces is usually considered. Forces acting on an object are balanced if the object is stationary, or if it is moving at a constant velocity. The driving force causing water to flow (whether in a channel, rill, gully or overland) is the down-slope component of gravity. This acts on a given mass of water, causing it to deform (flow) and move in a downstream direction over the channel boundary (bed and banks). Opposing this movement are resisting forces. Resistance occurs because of friction between the flow and channel boundary. Also, the fluid itself resists deformation because of internal forces within the flow. As water moves down slope, it exerts a shearing force, or shear stress, on the channel boundary (shear stress is represented by the Greek letter tau,). The **bed shear stress (**t**0**) is expressed as a force per unit area of the bed (in N m-2) and increases with flow depth and channel steepness. This relationship is described by the du Boys equation (Box 6.1).

Channel parameters

In order to describe the flow of water in river channels it is necessary to define some basic channel parameters, most of which are illustrated in Figure 6.1. Channel size can be defined by its cross-section: a slice taken across the channel, perpendicular to the direction of flow. The area of the cross-section is given by the product of channel width and the mean flow depth. At a given cross-section, the cross-sectional area changes through time in response to fluctuations in discharge (defined in previous chapters). The maximum discharge that can be contained within the channel, before water starts to inundate the floodplain, is called the **bankfull discharge**. The width of the channel at bankfull discharge is called the **bankfull width**. It should be noted that there are several issues associated with the definition of bankfull discharge for many river systems (see Chapter 3, p. 32). The shape of a river channel affects its hydraulic efficiency, something that can be quantified by calculating the **hydraulic radius**. This is a measure of how much contact there is between the flow and channel boundary, and is calculated from:

$$\text{Hydraulic radius} = \frac{\text{Cross-sectional area } (m^2)}{\text{Wetted perimeter } (m)}$$

The **wetted perimeter** is the length of channel boundary that is in direct contact with the flow at a given cross-section. An example is provided in Figure 6.2, which shows two channel cross-sections. For the purposes of this illustration, it will be assumed that the only difference between them is their shape, channel A is wide and shallow, while channel B is narrow and deep. Both have the same cross-sectional area but the wetted perimeter is larger for channel A, resulting in a lower hydraulic radius. Assuming all else is equal, the loss of energy arising from friction with the bed and banks will be greater for channel A. Channel B is therefore more hydraulically efficient. For wider channels, the hydraulic radius is very similar to the flow depth.

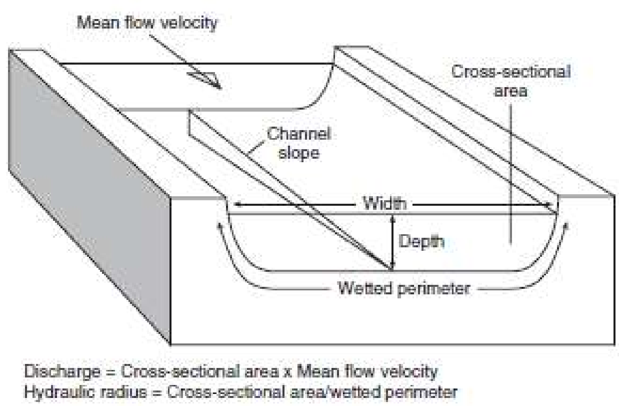


Figure 6.1 Basic channel parameters. After Summerfield (1991).

Channel slope is usually expressed as a gradient (difference in channel bed elevation along a given length of channel in meters divided by that length in metres). This is related to, but not necessarily the same as, the **water surface slope**, the downstream change in water surface elevation along the channel. Water surface slope is an important variable because it closely approximates the **energy slope** along a particular length of channel. As water flows through the channel, potential energy is converted to kinetic energy. This is in turn converted to heat energy, which is generated as a result of friction,2 and 'lost' from the channel. As a result there is a downstream reduction in the total energy 'possessed' by a given parcel of water. The steepness of the energy slope reflects the rate at which energy is being expended. Further detail on energy relationships is provided in Box 6.2.

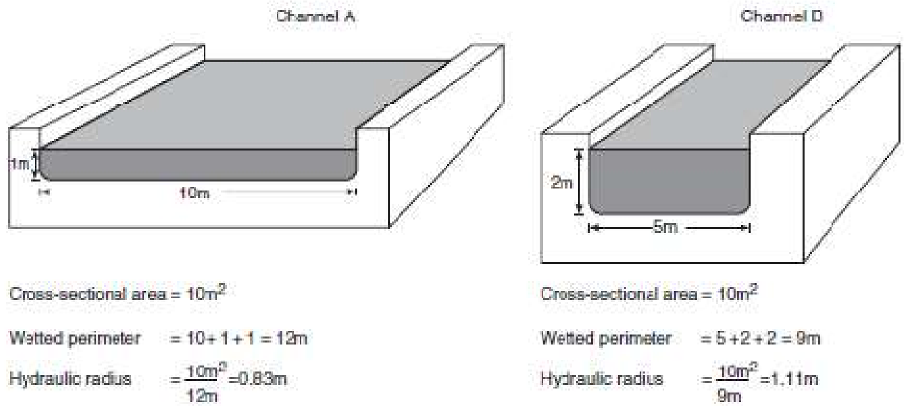


Figure 6.2 The effect of channel shape on hydraulic radius: wide, shallow channels are less hydraulically efficient than narrower, deeper channels. Adapted from Kay (1998).

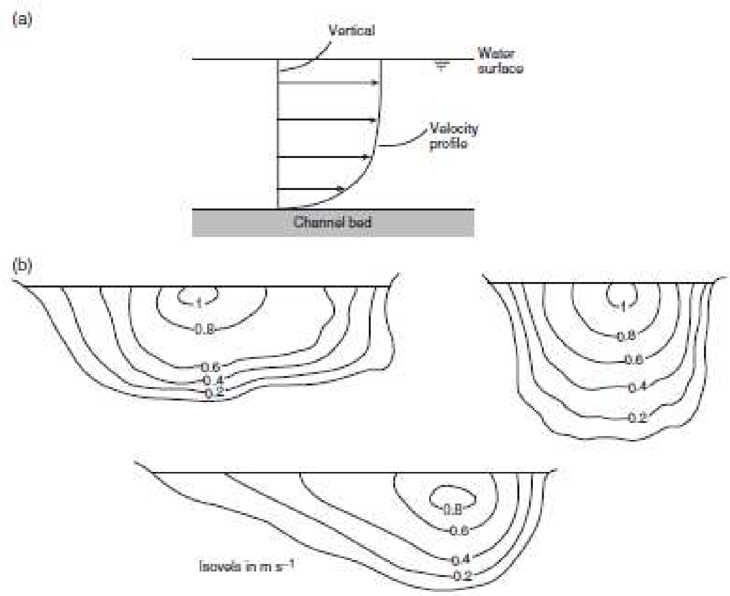


Figure 6.3 (a) Example of a vertical velocity profile, showing changes in time-averaged velocity with depth.
(b) Channel cross-sections illustrating the effect of channel shape on the velocity distribution. (b) is after Knighton (1998).

Flow velocity

Flow velocity varies over both space and time in natural channels. It is determined mainly by the channel slope, roughness and cross-sectional form (remember that channel depth and cross-sectional area change with discharge). If you have ever waded out into a stream, you will know that the flow velocity, like the depth of flow, tends to increase as you move out into the channel. This is because of friction between the flow and the channel boundary, which is greatest near the bed and banks. Together with the effects of turbulence, the effects of this frictional resistance create variations in velocity distributions that are seen at different spatial and temporal scales. These are briefly discussed below.

● ***Variations with time.*** At any given point within the flow, the velocity fluctuates rapidly because of the effects of turbulence. This means that instantaneous velocities at a specific location can be much higher or lower than the **time-averaged velocity** that is recorded by a flow meter (described in Box 3.1). Over periods of days, weeks or months, variations in velocity are also seen at the channel scale in response to discharge fluctuations.

● ***Variations with depth.*** These can be seen from measurements of time-averaged velocity made at different vertical heights above the channel bed (imagine a vertical line stretching upwards from a specific point on the channel bed). An example of a **vertical velocity profile** is shown in Figure 6.3a. At the bed itself, the velocity is zero, but increases with vertical distance above the bed. The actual *rate* of increase, or **velocity gradient**, is greatest close to the bed, levelling off further away from the bed. The vertical velocity gradient at any point determines the shear stress exerted on the bed at that point.

● ***Variations across the cross-section.*** Velocity profiles for three different cross-sections are shown in Figure 6.3b, where **isovels** – lines of equal velocity – have been plotted. As can be seen, the fastest flow occurs towards the centre of the channel. At this cross-sectional scale, the average flow velocity can be calculated by making a number of measurements of velocity across the channel and at different depths. A description of how to do this is sprovided by Goudie (1981).

● ***Downstream variations.*** Although there is typically a decrease in channel slope along the length of a channel, the velocity generally shows little change or increases slightly. This is because the decrease in slope is often compensated for by a downstream decrease in channel roughness and an increase in hydraulic efficiency. The effects of slope, roughness and hydraulic efficiency are discussed later in this chapter. Downstream changes in velocity will be considered in Chapter 8.

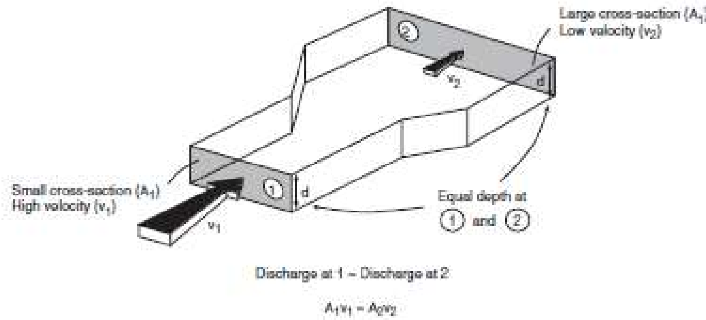


Figure 6.4 Illustration of the concept of flow continuity.

The concept of flow continuity

A casual observer walking alongside a natural stream channel might notice that the deeper sections are relatively slow flowing, while the shallow sections are relatively fast-flowing. The reason for this is that – assuming no tributaries join the channel and there are no significant interactions with groundwater – the same volume of water has to travel through each section in a given time. If this did not happen, the flow would start building up in some parts of the channel. Other parts would run dry as water flowed downstream faster than it was supplied from upstream.

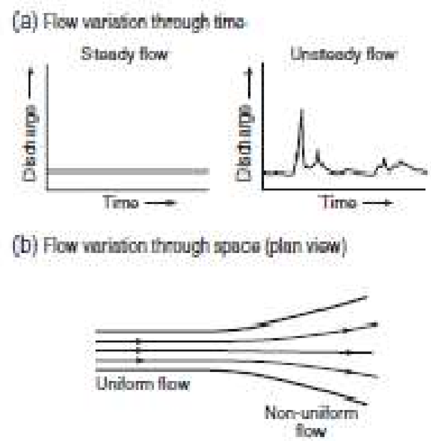


Figure 6.5 (a) Flow variation through time: steady and unsteady flow. (b) Flow variation through space: uniform flow and non-uniform flow.

not change either. Therefore the mass of water passing (1) is equal to the mass of water passing (2). If this was not the case, water would be spontaneously created or destroyed somewhere along the channel.

The concept of flow continuity is illustrated in Figure 6.4, which shows a length of artificial, watertight channel with a constant discharge entering at the upstream end. The discharge passing through cross-section (1) at the upstream end of the reach is equal to the discharge flowing through cross-section (2) at the downstream end. Since discharge is the product of cross-sectional area and flow velocity, it follows that: Discharge at (1) □□Discharge at (2) where *A1* and *A2* □□cross-sectional area at (1) and crosssectional area at (2); *v1* and *v2* □□mean flow velocity at (1) and (2).This is known as the **volumetric continuity equation**. From the example given in Figure 6.4 you can see that although the upstream cross-sectional area at (1) is smaller than the downstream one at (2), the upstream velocity is faster, so the same volume of water passes through the smaller cross-section over a given period of time.3 Following on from this is the concept of **conservation of mass**. The mass of a given volume of water can be calculated by multiplying that volume by its density. According to the volumetric continuity equation above, the volume of flow does not change and, since water cannot be compressed, its density (1 kg per litre) does not change either. Therefore the mass of water passing (1) is equal to the mass of water passing (2). If this was not the case, water would be spontaneously created or destroyed somewhere along the channel.

Variations through time: steady and unsteady flow

Steady and unsteady flows are classifications of flow variations through time. In the example above, it was assumed that the discharge entering the reach did not change through time, something that is called **steady flow**. In natural channels, the flow is usually **unsteady**, varying through time as the drainage basin responds to inputs of precipitation (Figure 6.5a).

Variations through space: uniform and non-uniform flow

Flows can also be classified according to variations over space. In a channel reach with a constant slope and cross-sectional shape, there will be no variation in either depth or velocity along the reach. This is called **uniform flow** and occurs in the upstream segment of the channel illustrated in Figure 6.5(b). The uniformity of the flow is indicated by streamlines – lines indicating the mean direction of individual ‘parcels’ of flow – which are parallel. Although most hydraulic equations are based on the assumption that the flow is uniform, this is rarely the case for natural channels, where the shape and

dimensions of the channel vary in a downstream direction. There are also bends and obstacles to flow such as constrictions and channel bars. The flow expands into wider sections and becomes concentrated where there are constrictions. This means that the streamlines are no longer parallel, and the flow is described as **non-uniform** (Figure 6.5b). It is only under uniform flow conditions that the channel bed slope, water surface slope and energy slope are the same. There are two types of non-uniform flow. **Gradually varied** flow is illustrated in Figure 6.5(b) and reflects changes that occur over distances of tens of metres or more. **Rapidly varied** flow is associated with sudden changes in channel width, depth or alignment. In these situations, the streamlines cannot follow the line of the channel and something called flow separation occurs. Hydraulic jumps and drops are also associated with rapidly varied flow. These types of flow behaviour will be discussed in the section after next.

FLOW RESISTENCE

A surprising amount of energy has to be used by flowing water in order to overcome flow resistance. It has been estimated that as much as 95–97 per cent of the total energy of a river is expended in this way (Morisawa, 1968). **Flow resistance formulae** express the relationship between flow velocity, channel slope, roughness and cross-sectional shape. Velocity increases with channel slope, but decreases with increasing boundary roughness. For example, a concrete-lined channel offers much less frictional resistance than a rocky, boulder-strewn channel. The hydraulic radius is also significant, since this determines the area of contact between the flow and boundary. Roughness is difficult to measure directly, so resistance formulae include an empirically-derived **friction coefficient** (examples are given in Box 6.3).

Channel resistance

At the valley scale, flow resistance increases when the channel comes into contact with the valley margins. This occurs in confined valley settings and where there are changes in valley alignment. The three-dimensional shape of the channel is also influential, since resistance is increased by irregularities in the banks, downstream changes in cross-section, and where the flow moves around bends. Bedrock-influenced channels can be highly irregular in form, with large variations in slope, width and channel cross-section. The high resistance of such channels is further increased by features such as cascades, vertical steps and potholes which increase form resistance (see below). In a detailed investigation of variations in total flow resistance for different channel types along the Sabie River in South Africa, Heritage *et al*. (2004) reported extreme values of total flow resistance for bedrock influenced channel reaches during low flows. These values were calculated for mixed anabranching channel sections, where the flow is divided into a number of separate bedrock-dominated distributary channels under low flow conditions (the distributaries are separated by bedrock core bars that are overlain by cohesive sediment and vegetation). At low discharges, the flow in each distributary is very shallow and the highly fissured bedrock pavement means that the wetted perimeter is very large and tortuous (a large wetted perimeter means a smaller hydraulic conductivity and greater flow resistance). Numerous pools and rapids form within the fissures, with steep water surface slopes and very high rates of energy dissipation. Added to this are the effects of numerous boulders, which create obstacles to the flow. As discharge increases, a decrease in resistance is seen as these features become increasingly submerged by the flow. During flood flows, the vegetated bars separating the distributary channels become inundated, with an increase in resistance that is attributed to the increased resistance of the vegetation (Heritage *et al*., 2004).

Boundary resistance

There are two components of boundary resistance. The first of these, **grain roughness**, relates to the effects of the individual grains making up the channel boundary. **Form roughness** refers to features such as ripples and dunes, which are created when certain alluvial substrates are moulded by the flow.

Grain roughness In general terms, flow resistance increases with the diameter of individual grains. However, an important factor is the depth of flow relative to the size of the particles. This can be expressed in terms of a ratio: where *d* is the flow depth and *D* is a characteristic grain size index; the median size of the bed sediment is often used. This ratio is used in many process-based equations in fluvial geomorphology and acts as a very significant control on the overall resistance in a channel (Robert, 2003). Bathurst (1993) compares the ratio of flow depth to characteristic grain size for different channel types along an idealised channel system. For a sand-bed channel, the flow depth may be over a thousand times greater than the diameter of the individual sand grains (2 mm or less). For gravel-bed channels, the ratio may be between 5 and 100, depending on the dominant grain size, which can range from cobbles (up to 250 mm) down to fine gravels (10 mm). In boulder bed channels, where most particles have diameters of 250 mm or more, the particles may project through the whole depth of flow, with a *d/D* ratio of less than 1. Where the stream bed consists of gravel or cobbles, grain roughness can be the dominant component of flow resistance (Knighton, 1998).

However, the effect of grain roughness is often drowned out' as the depth increases. Grain size, and the spacing of individual grains, can also have a significant influence on the structure of turbulent flows. These effects will be discussed in the next section (see section on 'hydraulically rough and smooth surfaces'). Form roughness In sand-bed channels, it is possible for a wide range of flows to shape the channel bed. At different flow intensities, a sequence of **bedforms** develops (the formation of which will be discussed in Chapter 7). These include dunes, which are scaled to the depth of flow in the channel. Bedforms increase turbulence and can cause flow separation, leading to significant energy losses at high flows. Varying levels of resistance are associated with different types of bedforms (Simons and Richardson, 1966) and in sand bed streams the presence of these forms often exceeds grain roughness in importance (Knighton, 1998).

In gravel-bed rivers, longitudinal variations in channel slope and bed roughness are often associated with periodic features called **riffles and pools**. Increased flow resistance occurs mainly as a result of ponding upstream from the shallower riffles (Hey, 1988). These are interspersed by deeper, slower moving pools, with a spacing of between five and seven times the channel width. Channel bars also increase flow resistance, particularly in braided channels. Even at higher flow stages, bars can account for between 50 and 60 per cent of total flow resistance (Prestegaard, 1983). Micro-scale variations in the bed topography of many gravel-bed channels are associated with smaller features called cluster bedforms. These consist of a single protruding obstacle, such as a large pebble, with associated accumulations of finer material immediately upstream and downstream (see Chapter 7). Cluster bedforms have a significant effect on shear stress distributions and flow resistance (Lawless and Robert, 2001).

In steep, rocky channels, sequences of **steps and pools** may form. These are associated with very high rates of energy expenditure, particularly at low flows when considerable energy has to be dissipated in hydraulic jumps and pools (Bathurst, 1993). The extreme resistance reported by Heritage *et al*. (2004) for the Sabie River was associated with bedrock-influenced rapids and cataracts. This morphology is highly irregular in form, with very high energy dissipation caused by hydraulic jumps, constrictions and other disturbances to the flow.

Other controls on flow resistance

Riparian and in-channel vegetation increases flow resistance. This varies with stage, because vegetation that is upright at low flows may become flattened at higher flows. Seasonal effects are also seen when vegetation dies back during the winter months. Patchy growth can lead to considerable variations in resistance across the channel bed. In some channels woody debris builds up to create additional resistance. Sediment transport may also be of some significance. A high suspended load increases fluid viscosity, reducing turbulence and, in turn, flow resistance (Knighton, 1998). Several studies have investigated the dependence of flow resistance on bedload transport for the coarse bed materials typical of mountain rivers. However these effects appear to be small in relation to other controls (Bathurst, 1993).

UNIT – 11: STREAM POWER AND ENERGY

THE CONCEPT OF STREAM POWER

Sediment entrainment, transport and deposition all involve the interaction of forces. **Work** is carried out when a force moves an object, the amount of work being defined by the size of the force and the distance over which the object moves. Work is involved in moving water through the channel and in eroding and transporting sediment. **Energy** is the capacity or ability to do work, and the same units, joules (J), are used for both.1 **Power** defines the rate at which work is done and is measured in watts (W), or joules per second. The concept of power can be illustrated by considering the transport of a piece of gravel between two points. This could be accomplished in a short period of time by a large force (high power), or over a longer period by a smaller force (low power). Although the same amount of work is involved in each case, it is carried out at different rates. **Stream power** is measured in watts per unit length of stream channel, usually W m−1. Stream power determines the **capacity** of a given flow to transport sediment. This is the maximum volume of sediment that can be transported past a given point per unit time. The available stream power is related to the water surface slope (*S*) and discharge (*Q*) of the channel. It is also affected by the gravitational constant, *g*, and the mass density of the fluid (1,000 kg m2 for water), which is represented by the Greek letter rho (□). These are combined in the equation below, where stream power is represented by the Greek capital letter omega ( ): Stream power is often defined in terms of the **specific stream power**, or stream power per unit area of the bed (per m2). Specific stream power (lower-case omega, □) is calculated by dividing the stream power per metre length of channel by the width of the channel. Where *W* is channel width. This is useful for making comparisons between rivers, or different reaches of the same river, because it reduces the scale effects of large and small channels. For British rivers, the specific stream power ranges from less than 10 W m2 for lowland channels in parts of the south-east, to 1,000 W m2 for rivers in the north and west, which drain steep upland areas with high rainfall (Ferguson, 1981). An example application of the stream power equations is shown in Box 7.1.

Box 7.1

EXAMPLE APPLICATION OF THE STREAM POWER EQUATIONS

In order to calculate the stream power per unit length for a reach of river channel you would need to measure the slope of the channel and the discharge. Say the channel slope, S, is 0.01 m m^{-1} and the discharge, Q, is 4 m^3s. We know the density of water, ρ (1000 kgm^3), and the gravitational constant, g (9.8 m s^{-2}), so:

$$\text{Stream power, } \Omega = \rho g Q S$$

$$= 1{,}000 \text{ kg m}^{-3} \times 9.8 \text{ m s}^{-2} \times 4.0 \text{ m}^3 \text{ s}^{-1} \times 0.01 \text{ m m}^{-1}$$

$$= 392 \text{ W m}^{-1} \text{ or } 0.39 \text{ kW m}^{-1}$$

This is the rate at which potential energy is converted to kinetic and heat energy along each metre of channel length. To calculate the stream power per unit area, it is necessary to measure channel width, W, and divide stream power by this.
Say channel width, W, is 3.7 m:

$$\text{Specific stream power, } \omega = \frac{\Omega}{W}$$

$$= 105 \text{ W m}^{-2}$$

In order to compare the stream power of two different channels, or different reaches of the same channel, it is important to take measurements when the flow conditions are comparable, as it would be fairly meaningless to compare low flow conditions in one channel with bankfull flow in another. For this reason specific stream powers are usually compared for bankfull flows.

Specific stream power can be related to bed shear stress (□0) and (cross-sectional) average flow velocity ( ): This means that the power per unit area of the bed is equal to the product of the average bed shear stress and the average flow velocity. Flow **competence** is the ability of a given flow to entrain sediment of a certain size and increases with bed shear stress.

UNIT – 12: MECHANICS OF FLUVIAL EROSION: THRESHOLD OF EROSION, ENTRAINMENT AND BED EROSION, HEADWARD EROSION AND CHANNEL LENGTHENING, BANK EROSION BY FLUVIAL PROCESSES

PROCESSES OF EROSION IN BEDROCK CHANNELS

The morphology of bedrock channels is mainly influenced by processes of erosion because the supply of sediment is often limited. Three types of erosion are significant: block quarrying, abrasion and corrosion. **Block quarrying** is the dominant process (Hancock *et al*., 1998) and involves the removal of blocks of rock from the bed of the channel by drag and lift forces. The size of the quarried blocks can be considerable. Tinkler (1993) reports blocks of sandstone 1.2 m × 1.45 m × 0.11 m and 1.0 m × 0.5 m × 0.05 m being removed from the bed of Twenty Mile Creek, Niagara Peninsula, Ontario, during normal winter flows, when the flow depth was less than 0.4 m. Before blocks can be entrained by the flow, a certain amount of 'preparation' is required to loosen them. Subaerial weathering and other weakening processes play an important role in this. Weakening processes described by Hancock *et al*. (1998) include the bashing of exposed slabs by particles carried in the load and a previously undocumented process termed 'wedging', which leads to the enlargement of cracks in the bedrock substrate. This is thought to occur when small bedload particles are able to enter cracks that are momentarily widened by fluid forces. The particles then become very firmly lodged and prevent the crack from narrowing again. As time progresses, further widening of the crack can be sustained as larger particles fall into it, and may ultimately lead to block detachment. Under conditions of very high flow velocity, sudden changes in pressure can generate shock waves that weaken the bed by the process of **cavitation**.

This effect is caused by the sudden collapse of vapour pockets within the flow (Knighton, 1998). **Abrasion** is the process by which the channel boundary is scratched, ground and polished by particles carried in the flow. Erosion is often concentrated where there are weaknesses and irregularities in the rock bed, which allow abrasion to take place at an accelerated rate. This can lead to the development of **potholes**, deep circular scour features that often form in bedrock reaches. Once a pothole starts to develop, the flow is affected, focusing further erosion.

Any coarse material that collects in the pothole is swirled around by the flow, deepening and enlarging it, and literally drills down into the channel bed. Over time potholes may coalesce, leading to a lowering of the bed elevation. Plate 7.1 shows how potholes have contributed to bed lowering near the site of a waterfall. Scouring by finer material carried by the flow, such as sand, leads to the development of **sculpted forms**. These include flutes and ripple-like features, which reflect structures within the flow (Plate 7.2). These are commonly observed on the crests of large boulders and other protrusions into the flow, where flow separation takes place and fine sediment is decoupled from the flow (Hancock *et al*., 1998). The rock boundary may also be polished by fine material carried in suspension.

Bedrock channels formed in soluble rock are also susceptible to erosion by **corrosion**, especially where the presence of joints and bedding planes allows solutional enlargement. Solutional features such as scallops may also be seen. These spoon-shaped hollows often cover the walls of cave streamways. Their length is related to the formative flow velocity, ranging from a few millimetres (relatively fast flow) to several metres (relatively slow). Although the actual processes of erosion operate at a small scale, their effects can be seen over scales ranging from millimetres to kilometres. There are several controls on rates of erosion, which influence the processes described above. These include micro-scale (millimetres to centimetres) variations in the rock structure, the larger scale effects of bedding, joints and fractures, and basin-scale influences such as regional geology and base level history (Wohl, 1998).

BANK EROSION IN ALLUVIAL CHANNELS

Processes of bank erosion are important in the development and evolution of different channel forms, while the migration of river channels across their floodplains involves a combination of bank erosion and deposition. Bank erosion can also create management problems when bridges, buildings and roads are undermined or destroyed. Large volumes of sediment can be generated, leading to problems of aggradation further downstream. Land disputes may also arise where boundaries lie along actively migrating river channels. Rather than being a process in itself, bank erosion is brought about by a number of different processes which can be considered in three groups:

1 *Pre-weakening processes* such as repeated cycles of wetting and drying, which 'prepare' the bank for erosion.

2 *Fluvial processes*, where individual particles and aggregates are removed by direct entrainment.

3 *Processes of mass failure*, which include the collapse, slumping or sliding of bank material into the channel.

Bank material that has been detached remains at the base of the bank until it is broken down *in-situ* or entrained and transported downstream. A balance exists between the rate of sediment accumulation and its rate of removal, which acts as an important control on rates of bank erosion (Carson and Kirkby, 1972). If material accumulates at the base of the bank at a faster rate than it is removed then, to a certain extent, the bank is protected from further erosion. When the opposite situation applies, with bank material being removed faster than it accumulates, bank erosion will continue, sometimes at an increased rate. A third possibility is that rates of supply are the same as rates of removal. The relative rates of accumulation and removal are dependent on the available stream power and the controls on bank erosion discussed below.

Bank materials and weakening processes

The moisture content of the bank is significant, particularly for cohesive bank materials whose strength varies with the level of saturation. A certain amount of water is held in the pores, against the force of gravity, by matric suction forces. These result from surface tension effects, and a negative pore water pressure (less than atmospheric) develops when the soil is not completely saturated. As the soil dries, the strength of the matric suction forces increases as all but the smallest pores are emptied. These forces can be considerable and several authors have observed an increase in the resistance of the bank material to erosion at high matric suctions. However, it has also been suggested that desiccation can lead to higher rates of bank retreat, because the shrinking of clay particles causes cracking and shedding of loose material at the bank surface.

The process of **slaking** occurs when banks are rapidly immersed by floodwaters and air becomes trapped and compressed within the pores. The resultant pressure causes material to become dislodged (Thorne and Osman, 1988). At high flows, banks may become saturated with water from the channel. Saturation also occurs when there is a rise in the water table or during prolonged rainfall. Under these conditions a positive pore water pressure exists between the grains. This weakens the cohesive forces, acting as a lubricant and reducing inter-granular friction. During cold conditions, the growth of lenses, wedges, and crystals of ice can significantly reduce resistance to erosion, especially where freeze–thaw cycles occur. In temperate regions, the growth of ice needles occurs during moderately sub-zero temperatures. These are elongated crystals of ice that start to grow as the temperature of the air in contact with the bank decreases, growing in the direction of cooling (into the bank). The crystals often lift and incorporate material which then moves downslope or remains as a 'sediment drape' when the ice melts (Lawler, 1988). In colder regions, where rivers freeze over in winter, cantilevers of ice can cause significant damage (Church and Miles, 1982). Where permafrost exists, thermoerosion niches are cut into frozen banks by the relatively warm water in the channel.

While not a process in itself, the presence of vegetation influences the resistance to bank erosion in various ways. Root networks are particularly important and vegetated banks tend to have a more open structure and be better drained. Vegetation also acts to bind the soil together and increase the shear strength of the bank material. Unlike soil, roots have a very high tensile strength, which means that they are able to resist tension (stretching forces).

Bank erosion by fluvial processes

For any given situation the relative importance of direct entrainment and mass failure is mainly determined by the composition of the bank, although other factors can also be important. Banks composed of sand and coarser particles are non-cohesive and this material is usually detached grain by grain. Although cohesive forces do not exist between the particles, movement is resisted by inter-particle friction and the packing structures holding the grain in place. However, the selective entrainment of finer sands and gravels often leads to a weakening of the overall structure, which may lead to collapse. In the case of cohesive banks, it tends to be aggregates and crumbs that are detached rather than individual particles. The weakening processes described above are of great importance in assisting fluid forces to detach and entrain aggregates. Once entrained into the main flow, aggregates tend to disintegrate fairly rapidly.

Bank failure mechanisms

Bank failure occurs when bank material becomes unstable and falls or slides to the base of the bank. There are several types of failure, and different failure mechanisms are observed for cohesive and non-cohesive bank materials. Also important are bank height, bank angle, moisture content and the effects of vegetation. There are some similarities between bank failure mechanisms and the processes of mass wasting discussed in Chapter 4 (pp. 39–42). The stability of banks is determined by the balance between the shear stress exerted by the down-slope component of gravity (driving force) and shear strength of the bank material (resisting force). In cohesive banks, failure occurs across a **failure plane**, the surface within the bank across which shear stress exceeds shear strength. Failure planes can be almost planar (flat) or curved. One of the most common types of failure is illustrated in Figure 7.1(a) and occurs where banks are low, steep and composed of cohesive material. Typically the failure surface is almost planar and vertical, parallel to the bank surface (Plate 7.3, see also Colour Plate 3). Where bank angles are less steep, the failure plane is usually curved and located deep within the bank (Figure 7.1b). Cohesive banks are often most susceptible to failure after a flood wave has passed, when the saturated banks are no longer supported by the pressure of flow in the channel. Non-cohesive banks tend to fail along shallow slip surfaces (Figure 7.1c). Mixed banks are common, typically with fine cohesive sediment overlying non-cohesive material (Plate 7.4). Undercutting of the non-cohesive material by fluvial processes leads to instability of the overlying material. This can cause various types of bank instability, including the cantilever failure illustrated in Figure 7.1(d).

UNIT – 13: RIVER TRANSPORTATION: PROCESS OF PARTICLE ENTRAINMENT, INITIATION OF SEDIMENT TRANSPORT, SELECTIVE THEORIES OF SEDIMENT TRANSPORT, THEORIES OF EQUAL MOBILITY TRANSPORT, TRANSPORT OF BED LOAD AND SUSPENDED LOAD

SEDIMENT ENTRAINMENT AND TRANSPORT

The process of particle entrainment

Whether or not a given particle is set in motion depends on the balance between the forces driving and resisting its movement. These are illustrated in Figure 7.2. The resisting force is the immersed weight of the particle (in this simple example the effects of neighbouring grains will be ignored). The driving force is provided by the combined effect of two **fluid forces** exerted on the particle by the flow: a drag force and a lift force. The fluid **drag force** acts in the same direction as the flow and can be thought of as the 'force of the flow' that is felt when you wade out into the current of a stream. It comes about because the pressure exerted on an object by the flow is greater on its upstream side than on its more sheltered downstream side. The second fluid force, the **lift force**, acts vertically upwards and is caused by a pressure difference above and below the particle. Water flowing over the particle has to move faster. According to the results in a decrease in pressure above the particle, while the pressure below it stays the same. This difference in pressure generates lift. In theory, if this force exceeds the gravitational force, the particle will be lifted from the bed. In practice, the presence of other particles complicates matters considerably. Sediment transported as bedload is generally gravelsize and larger, although coarse sands may also form part, or all, of the bedload component. Finer bedload, which is too coarse to be transported in suspension, is moved along the bed in a series of short jumps by **saltation**. Saltating grains are lifted from the bed at a relatively steep angle by the combined forces of lift and drag. As a grain moves upwards into the flow, the lift force decreases and it starts to fall back towards the bed. The falling grain is carried downstream by the drag force, following a shallow trajectory towards the bed. Larger particles, which cannot be lifted, are rolled or dragged along the bed. This movement is usually sporadic because of variations in bed shear stress. In addition, particles tend to become lodged behind other particles or obstacles on the bed. The weight of smaller particles carried in suspension is supported by turbulence. Descending saltating grains may also be temporality lifted upwards by turbulent movements. This is called incipient suspension.

Plate 7.3 Slab failures in cohesive banks, River Ure, North Yorkshire, England.

Plate 7.4 Mixed bank, where selective entrainment of the underlying non-cohesive material has led to collapse of the cohesive layer above. Gunnerside Gill, North Yorkshire, England.

Size-selective theories of sediment transport

A considerable amount of research has focused on deriving critical flow or **entrainment thresholds** from easily measured flow parameters. There are various practical reasons why we might want to know the flow conditions that will move particles of a certain size. These include the planning of reservoir releases to flush out fine sediment from fish spawning grounds (without removing the gravel), or determining when structures such as bridge piers are at risk of being undermined by erosion.

The threshold conditions for the entrainment of particles of a given size can be defined according to a critical mean flow velocity (i.e. cross-sectional average) or a critical bed shear stress. Using the mean flow velocity is an indirect method, since it is actually the hydraulic conditions near the bed of the channel that are significant. However, both relationships show similar basic trends. An explanation will first be given in terms of a critical mean flow velocity since this relationship is conceptually easier to understand.

The critical mean flow velocity curves shown in Figure 7.3(a) were derived from a large amount of experimental data accumulated by Filip Hjulstrøm in the 1930s. They show the entrainment and fall (or settling) velocities for particles of different sizes, from fine clay to coarse gravel and small boulders. Note that a logarithmic scale is used on both axes to cover the wide range of particle sizes and the corresponding range of flow velocities. The upper curve on the graph shows the **entrainment velocity** required to set different particle sizes in motion. Sand grains, with a diameter of between 0.2 mm and 0.7 mm, are the easiest to entrain. In the case of larger particles, which have a greater immersed weight, the entrainment velocity increases with particle size as might be expected. However, the relationship is rather different for particles smaller than 0.2 mm, since the entrainment velocity actually *increases* as the particle size decreases from fine sand to silt and clay. Reasons for this include the fact that these small particles tend to be partly or wholly enclosed within the laminar sublayer during most flows. Drag forces are lower within this layer, and particles are not exposed to turbulent lift forces. In addition, the cohesive forces between clay particles further increases the force required to set them in motion. An alternative approach, which is more relevant to modern sediment transport theory, was devised by the American engineer Albert Shields in 1936. This defines the critical bed shear stress necessary to set particles of a given size in motion. The critical bed shear stress is actually defined in a dimensionless form. The dimensionless critical bed shear stress is often referred to as the **Shields parameter**. It appears in a number of sediment transport equations and is represented by the Greek letter theta (□c – the subscript is short for 'critical'). The equation defining dimensionless bed shear stress is shown in Box 7.2. Critical bed shear stress increases with particle size but also depends on bed roughness. Shields related the dimensionless bed shear stress to the boundary Reynolds number (Re□). This was defined in Chapter 6 and represents the roughness of the channel bed (see 'hydraulically rough and smooth surfaces' on p. 84–5 and Box 6.4). The boundary Reynolds number is proportional to the ratio between grain size and laminar sublayer thickness. Figure 7.3(b) illustrates the Shields relationship, which is basically similar to the Hjulstrom relationship. Where Re□□is less than about 5, the grains are small enough to be fully submerged within the laminar sublayer. As these sheltered particles get smaller, the shear stress needed to entrain them increases. For hydraulically rough surfaces, the critical bed shear stress is independent of the boundary Reynolds number and the critical bed shear stress reaches a constant value of 0.06 (Richards, 1982). The lowest critical bed shear stress is associated with sand grains in the size range 0.2 mm to 0.7 mm (Knighton, 1998). It is important to note that the Hjulstrom and Shields experiments were carried out using well sorted bed sediment of a single size. This is not representative of the conditions on the bed of many channels, where there is a mixture of grain sizes. The arrangement of grains on the bed and the mixture of grain sizes is very significant, affecting both the entrainment of individual grains and overall transport rates.

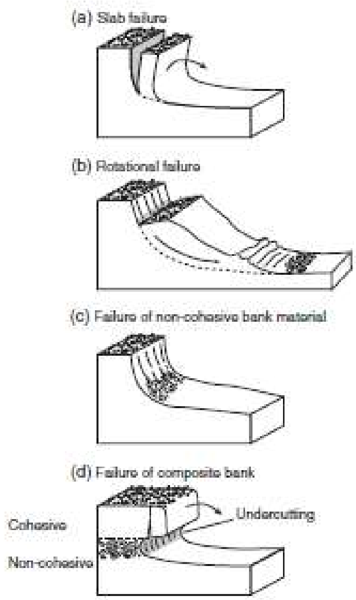


Figure 7.1 Bank erosion processes.

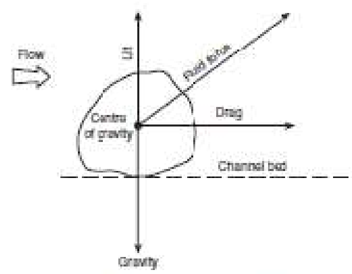


Figure 7.2 Driving and resisting forces acting on a grain of sediment resting on the bed of a channel. Adapted from Morisawa (1985).

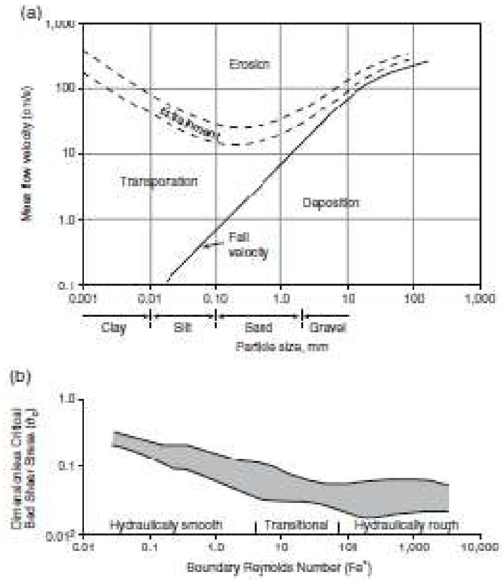


Figure 7.3 (a) The Hjulstrøm diagram, showing the mean flow velocity at which particles of a given size are entrained, transported and deposited. These curves are based on experimental data carried out for well sorted sediments of a uniform size. Adapted from Hjulstrøm (1935). (b) A Shields diagram showing the relationship between the critical dimensionless bed shear stress and the boundary Reynolds number. A large amount of scatter is associated with different experimental conditions and the problem of defining exactly when a threshold has been reached. Adapted from Miller *et al.* (1977) and Summerfield (1991).

Sediment transport in mixed beds

The mobility of individual particles is greatly affected by the size and arrangement of the particles surrounding them. In most natural channels the mixture of sediment sizes, and an irregular bed surface, makes the situation rather more complicated. Figure 7.4(a) shows how the arrangement of grains affects the 'rollability' of individual particles, which determines how easily a particle can be moved from its resting position (Pye, 1994). This can be defined in terms of a **friction angle**, which is greatest where small particles overlie larger ones, meaning that a greater force is required to pivot smaller particles away from the bed. Larger grains can also shelter smaller grains from flows that would otherwise be competent to entrain them (Figure 7.4b). The degree of **sorting** reflects the range of particle sizes in a particular sample of bed material. Well-sorted sediments have a narrow range of particle sizes, whereas poorly sorted material shows a much wider range. Box 7.3 explains how the distribution of grain sizes in a channel are described. In gravel-bed rivers, particles may also be wedged together in various types of packing arrangements which act to resist bed shear stresses and again make it much harder for individual grains to be entrained. Figure 7.4(c) shows how coarse gravels and pebbles can be deposited in an **imbricated** structure, inclined in the direction of the flow that deposited them, with the long axis of each particle overlapping the next.

Armour layers

In gravel-bed rivers the development of an **armour layer** has a very significant impact on rates of bedload transport. Armouring is illustrated in Figure 7.4(d) and develops during frequent low magnitude flow events that are only competent to entrain the smaller particles. As a result, fine sediment is removed from the bed leaving a layer of coarse sediment, usually about one particle diameter in thickness. This armour layer protects the finer material beneath from subsequent high flows. Once a bed is armoured, a much higher critical threshold is required to break it up. Bathurst (1987a) defined **'twophase' flow** for armoured channels. During phase 1 flow an armour layer is present and rates of bedload transport are low (although finer sediment can still be supplied from further upstream). Once the armour layer breaks up phase 2 transport takes place, with a dramatic increase in transport rates as the finer sediment becomes available. This can lead to complex variations in bedload transport through time. For example, where two high magnitude flow events occur in close succession, the initial rate of transport is often much higher for the second event than for the first, which breaks up the armour layer. Research has shown that ephemeral channels do not tend to develop an armour layer because, in the absence of low flows, there is no mechanism for removing fine sediment to create the armour layer. This could mean that rates of bedload transport are greater for ephemeral channels than for channels in humid settings (Nanson *et al*., 2002).

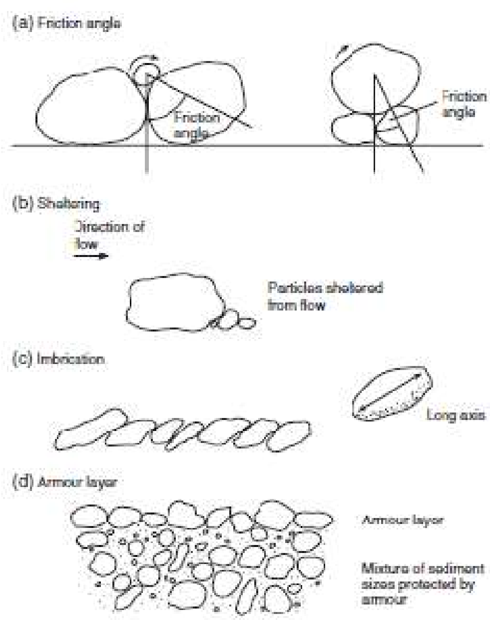


Figure 7.4 Bed sediment characteristics. (a) Friction angle. (b) Sheltering. (c) Imbrication. (d) Armour layer.

The theory of equal mobility transport

On a level bed with a uniform sediment size, all the particles might be expected to begin moving under approximately the same flow conditions (Reid *et al*., 1997). However, on mixed beds, the relative size of a given sediment particle determines its degree of exposure to the flow. As a result, larger particles shelter smaller particles, which then require a higher shear stress for entrainment than would otherwise be the case. In contrast, coarser grains are more easily entrained when surrounded by fine grains. This is because they are relatively more exposed to the forces of entrainment (Andrews, 1983). Particles of an intermediate size are relatively unaffected by the sheltering/hiding effects. Empirically, this 'reference size' has been shown to approximate the median size ($D50$) (Bathurst, 1987b). On the basis of field data, Parker *et al*. (1982) introduced a theory of **equal mobility** for channel beds composed of a mixture of sediment sizes. This states that the threshold condition for each size fraction is not dependent on the grain size. In other words, the movement of particles of different sizes can be initiated under similar critical flow conditions. The theory of equal mobility transport therefore challenges the size-selective transport theory of Shields (1936). However, any deviation away from an equal mobility condition represents some degree of size-selective transport. Under conditions of equal mobility, the bedload transport rate could be calculated from a single representative grain diameter such as the median size, $D50$ (Parker *et al*., 1982).

Equal mobility transport is the subject of some debate, however. Field investigations into the occurrence of equal-mobility transport have mainly been carried out in gravel-bed channels, where the largest grains are cobblesize or smaller (for example Andrews, 1983; Ashworth and Ferguson, 1989). Measurements made in rivers during steady uniform flows have shown that the transport of mixed sediment is only weakly size selective at low shear stresses. At higher shear stresses, sediment transport approaches equal mobility (Parker *et al*., 1982; Andrews, 1983; Marion and Weirach, 2003). However, observations made over a wider range of flows (e.g. Ashworth and Ferguson, 1989; Wilcock, 1992) have emphasised the size-selective nature of gravel transport. Only during the highest flows does sediment transport approach equal mobility. For example, Wilcock (1992) observed a progressive shift away from unequal to equal mobility transport with increasing shear stress, although equal mobility was not observed until the shear stress was over twice the critical stress required to initiate motion. One of the biggest problems associated with these investigations is obtaining sufficient field data to include a representative range of flow conditions (Reid *et al*., 1997).

BEDLOAD TRANSPORT

Bedload transport does not necessarily take place all the time, and rates may approach zero during low flows. Even when transport is occurring, it is likely that only part of the bed will be mobile at any one time. Part of the reason for this is the uneven distribution of bed shear stresses, which is directly controlled by variations in turbulent fluctuations. Large differences are observed across small areas of the bed and over short periods of time. Sweep fluid motions, inrushes of high momentum fluid from the outer zone of the boundary layer (Chapter 6, p. 86), are particularly effective at entraining bedload particles (Robert, 2003). The ejection of low momentum fluid away from the bed also allows finer sediment to be lifted up away from the bed and into the turbulent profile, maintaining it in suspension. The availability of bed sediment has an important influence on overall rates of bedload transport in a given reach of channel, and many bedload-dominated channels are transport limited. This means that transport rates might be lower than expected at a particular flow because of a lack of available sediment. This 'lack' does not necessarily refer to the total volume of bed sediment in a reach, more relevant is the availability of sediment of a certain size or calibre. Thus a flow that is competent to transport only fine gravels will not be able to entrain the larger material in a boulder-bed stream, no matter how abundant this is. The supply of bedload can be especially limited in bedrock channels and the flow capacity often exceeds that required to transport the available load.

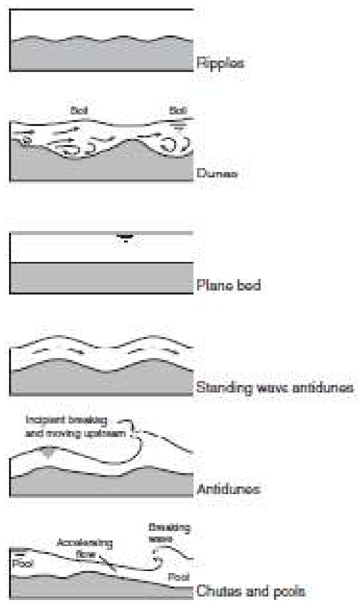


Figure 7.5 Bedform variation in sand-bed channels. Adapted from Simons and Richardson (1966).

s

Bedforms

Bedforms in sand-bed channels In sand-bed channels the sand grains can be transported at both high and low flows because of their low entrainment threshold. As a result the bed is easily shaped by flows to form periodic features known as bedforms. These have been intensively studied, both in the laboratory and in natural channels, and a recognisable sequence of bedforms develops in response to changing flow conditions (Figure 7.5). For the purpose of explanation, the starting point is assumed to be a plane bed, something that is rare in natural channels because even the smallest flows start to shape the bed. When water starts to flow over a flat bed the sand grains start to move, individually at first and then in patches, until periodic **ripples** develop, with crests perpendicular to the direction of flow. Field and laboratory research suggests that the wavelength, or spacing, of ripples is mainly dependent on particle size and is typically between 150 mm and 450 mm. As the flow intensity increases, ripples start to give way to **dunes**, larger features with rounded crests. Dunes are common in alluvial channels and are continuous along the bed for hundreds of kilometres in large rivers like the Mississippi and Niger. Dunes vary greatly in size, being scaled with the depth of flow and ranging from a few centimetres to a few metres in height. Dune wavelengths also vary, from tens of centimetres to more than a hundred metres in the largest rivers. Ripples and dunes migrate downstream over time, as the flow moves sand grains up the more gentle upstream slope towards the crest, from where sediment falls down the steeper downstream slope. A critical mechanism in this process is the deposition of coarse grains at the crest, where flow separation occurs. At higher flows, dunes become unstable and are 'washed out' because the flow velocity is too great to sustain deposition at the dune crest. Dunes then give way to a **plane bed**, but one that is rather different from the initial flat bed. Above the bed is a clearly defined zone of suspended sediment within which 'dust storm conditions' prevail (Leopold *et al*., 1964). This marks the transition to the **upper flow regime**, where the Froude number (ratio of inertial and gravitational forces) is greater than 1 and flow becomes supercritical. Upper flow regime bedforms include **standing wave antidunes**, where the sediment is moving but the waves themselves are stationary. This is because rates of deposition on the upstream side are matched by erosion on the downstream side. The position of standing waves is marked by waves at the water surface, the sand and water waves being in phase with each other. At higher flows sediment is thrown up from the downstream side of the bedforms at a faster rate than it can be replenished, which results in **antidunes**. These migrate *up*stream, while the sediment continues to move downstream.

Antidunes can be seen in the high-velocity gully flow shown in Colour Plate 15. At very high flows, a series of **chutes and pools** develop. Chutes have a near-plane bed and shooting flow, which enters downstream pools: deeper sections that are marked by hydraulic jumps. Bedforms in gravel and mixed sand–gravel channels Bed structures also form in gravel-bed channels and have been a focus of research over recent decades. **Pebble clusters** are commonly found in this type of channel (Figure 7.6) and form when a single large particle acts as an obstacle, protruding into the flow and encouraging the accumulation of coarse material on its upstream side. This upstream material may have an imbricated structure, increasing stability and requiring larger lift and drag forces to entrain the constituent particles. Finer particles are found on the downstream side of the obstacle, where shelter is provided from lift and drag forces. **Transverse ribs** are another type of gravel bedform and consist of regularly spaced ridges of coarser pebbles, cobbles or boulders that lie transverse to the flow. Like sand bedforms, these features affect flow resistance as well as rates of bedload transport.

Where the bed is composed of a mixture of sand and gravel, the different mobility of the constituent particles can lead to some interesting effects. For example, longitudinal ribbons of sand have been observed to travel downstream, snaking from side to side over immobile gravel beds. Bedload can also move in thin sheets as an elongated procession of sediment with a thickness of one to two grain diameters. The coarsest sediment accumulates at the leading edge and there is a progressive fining of sediment behind it. Bedload sheets appear to be fairly common in mixed channels and are related to rates of sediment supply, becoming less frequent and reduced in extent as supply rates are reduced (Dietrich *et al*., 1989).

Assessing rates of bedload transport

From the preceding discussion you will have some idea of just how complex and variable bedload transport is. There is a general paucity of data on rates of transport because the available techniques can be expensive and time-consuming to employ. These include the collection of bedload over a period of time using portable samplers or traps excavated in the bed. Another approach is to track the movement of individual particles. Further information on these techniques can be found in Box 7.4. A number of bedload transport formulae have also been developed, using field and laboratory data, to predict rates of transport from flow variables. An overview is provided in Box 7.5.

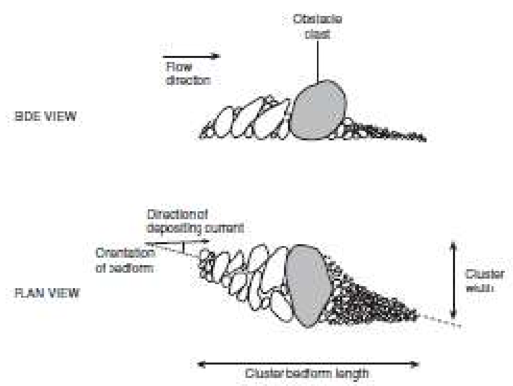


Figure 7.6 Idealised diagram of a cluster bedform. After Brayshaw (1984).

SUSPENDED LOAD TRANSPORT

Processes

Particles carried in suspension are kept aloft by turbulent eddies and will remain in suspension as long as their weight is supported by the upward component of turbulent eddies. In a fluid at rest, a suspended particle will fall through the fluid column. The rate of fall, or **fall velocity**, is a function of the density, size and shape of the particle. It is also determined by the viscosity and density of the transporting fluid. Since the falling particle displaces fluid, its movement is resisted by an equal and opposite fluid drag force. If sufficient depth is available, the falling particle will accelerate until it reaches a terminal velocity (defined for falling raindrops in Chapter 4). In channels, the fall velocity is further affected by flow turbulence and the interactions of surrounding particles (Chanson, 1999).

Considerable variation is seen between particles of different sizes. The fall velocity for the finest wash load component is very low, meaning that this sediment can be transported over considerable distances (Chapter 5). For example the terminal fall velocity of a silt grain (0.001 mm) is approximately 0.004 cm s-1, but increases to 34 cm s-1 for a 10 mm gravel particle (Chanson, 1999). Suspended sediment is transported by processes of **advection** and **turbulent diffusion**. Advection is the transport of sediment within the flow, where the sediment moves with the flow itself. Turbulent diffusion refers to the mixing of sediment through the depth profile by turbulent eddies. Within the depth profile, the greatest concentration of suspended sediment is found towards the bed of the channel. Although there is continuous movement of individual suspended grains, the overall concentration and average grain size generally decrease rapidly away from the bed. This is due to interaction between the fall velocity and the vertical component of flow associated with turbulent eddying (Knighton, 1998). The upward migration of sediment to zones of lower concentration is both an advective and a diffusion process.

A related process, which is called **convection**, involves the entrainment of sediment by large-scale vortices. For example, sediment is suspended in vortices generated as a result of flow separation in the toughs of ripples and dunes (Bridge, 2003). Large-scale vortices also occur where there are sudden drops in bed elevation, at hydraulic jumps and during overbank flows. Vortices created within the shear zone between the faster moving channel flow and slower flow on the floodplain result in lateral transfers of water. From Figure 6.10 you will see that there is a near surface flow out onto the floodplain and a return flow back towards the channel. Sediment carried in the flow is also transferred from channel to floodplain. Most of the coarser sediment is deposited near to the channel margins because of the rapid deceleration of flow and reduced turbulence (Bridge, 2003). This explains the origin of alluvial ridges, or natural levees, that are found along the margins of some channels. Finer suspended sediment, especially the wash load component, is carried out onto the floodplain.

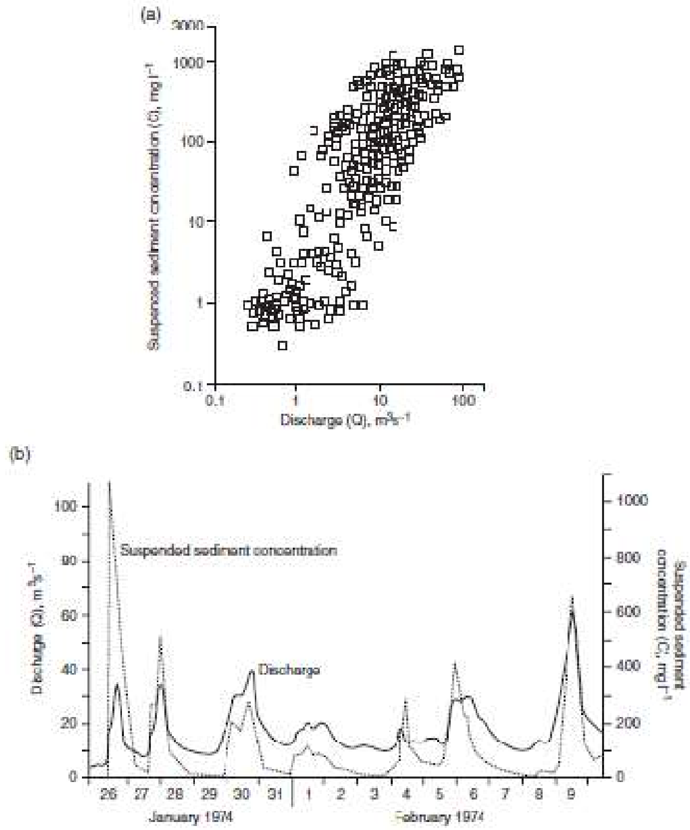


Figure 7.7 Characteristic of suspended sediment transport by rivers. (a) Plot of suspended sediment concentration versus water discharge for the River Creedy, Devon, England. (b) Typical record of the variation of suspended sediment concentration during a sequence of storm runoff events for the same river. After Reid *et al.* (1997).

Sediment supply and transport rates

The main sources of suspended sediment include material washed in from hillslope erosion and the release of fine material and aggregates from bank erosion. The supply of fine sediment is a major control on rates of suspended sediment transport. Most suspended transport, particularly the wash load, is supply limited. This means that the supply of fine sediment often has a greater influence on the sediment concentration than flow conditions in the channel. The rate of supply varies during individual events, between events, seasonally and annually. These variations are controlled by a number of variables, including antecedent conditions, rainfall intensity, hydrograph shape and vegetation growth. High discharges tend to be associated with greater concentrations of suspended sediment. This is because the supply is increased by storm-induced erosion of hillslopes and channel banks, and the release of fine sediment from storage. With all this in mind, it is hardly surprising that no simple relationship exists between suspended sediment concentration and flow discharge for a given crosssection. This is illustrated by the wide scatter of points in Figure 7.7(a), which shows this relationship for the River Creedy in Devon, England. Figure 7.7(b) shows flow and sediment hydrographs for the River Creedy. For the three events in January 1974 there is a decrease in the sediment load, even though the magnitude of the flow peak of the first two events is similar and the third event has a larger peak flow. This is an example of the '**exhaustion effect**' that can result when several events occur in close succession and the available sediment becomes depleted. The role of storage is important here because much of the sediment washed into river channels will have been released from temporary storage on hillslopes.

UNIT – 14: SEDIMENT LOAD: NATURE AND TYPES OF SEDIMENT LOAD, SOURCES OF SEDIMENT LOAD

WEATHERING

Weathering involves the physical breakdown and chemical decomposition of rocks and loose rock material *in situ* at the Earth's surface. It takes place when these parent materials are exposed to environmental conditions at the Earth's surface, where many rocks become unstable. This is because the conditions there may be very different to those under which they were formed, perhaps at high temperatures and pressures some depth below the surface. Weathering produces a primary source of sediment for removal by processes of erosion and mass wasting. It also provides mineral matter as a raw material in soil formation. Here we are mainly interested in weathering as a source of sediment, so the actual processes will not be considered in detail.

Physical weathering processes

Weathering processes may be classified as physical or chemical, although this is a theoretical distinction because in most cases physical and chemical weathering act together. Physical weathering involves the mechanical breakdown of rocks into smaller fragments without any change taking place in their chemical condition. Processes include breakdown due to stresses exerted by the growth of crystals of ice or salt in rock crevices; pressure release caused by the denudation and removal of layers of overlying rocks; and thermal expansion and contraction caused by heating and cooling due to wildfire or solar radiation. In mountainous regions, where the temperature regularly fluctuates above and below zero, **freeze– thaw action** is a very effective breakdown process. Water expands when it freezes and the repeated formation and melting of ice crystals can exert considerable stresses.

Chemical weathering processes

Chemical weathering brings about changes in the chemical, mineralogical and physical properties of rocks. Alteration of the constituent minerals often leads to a weakening of the overall structure. For example, many of the secondary minerals formed by the breakdown of primary minerals in rocks are less dense but take up a greater volume. The most important group of secondary minerals are the clay minerals.

Hydrolysis is one of the main chemical weathering processes. It can drastically modify and decompose susceptible primary minerals, and is significant in the conversion of parent material to clay. Hydrolysis involves a reaction between water and silicate minerals in rocks and soil in which mineral cations such as potassium, calcium and magnesium are replaced by hydrogen ions. This weakens the internal structure and leads to breakdown.

Carbonation dominates the weathering of calcareous rocks, such as limestone and dolomite. It is a complex process, brought about when carbon dioxide from the atmosphere is dissolved in rainwater to form a weak carbonic acid. This is able to attack calcareous rocks by forming water-soluble carbonates. Carbonation is a step in the complex weathering of many other minerals, such as the hydrolysis of feldspar (Huggett, 2003). Water is also involved in the more 'mechanical' processes of **solution**, where mineral salts dissolve in this very effective solvent, and **hydration**, where the water molecules are absorbed by certain minerals but do not alter the minerals chemically. Weathering reactions involving oxygen also occur. Iron-bearing minerals are susceptible to **oxidation** where oxygen dissolved in soil waters combines with minerals within the parent rock. This alters the crystal lattice and makes it more prone to further breakdown (Huggett, 2003). **Weathering products** The surface of exposed bedrock is often mantled by a layer of weathered material called **regolith**. This includes solid particles, ranging in size from large boulders to fine clay particles, material carried away in solution and colloids (microscopic particles). The thickness of the weathering mantle is determined mainly by the balance between rates of weathering, and the relative rate at which material is transferred away from the site by mass movement and agents of erosion (water, wind and ice).

Rates and controls

Rates and types of weathering are determined by four main groups of variables: climate, parent material, topography and organic activity. These operate over different scales. For example there is a reasonable correlation between the major climatic regions and different weathering zones, while the effects of topography and parent material tend to be more localised. The two climate-related parameters of greatest significance are temperature and precipitation, the latter being the primary control on water availability. Rates of chemical weathering increase with temperature, approximately doubling for every 10°C rise. Organic activity also increases with temperature and leads to the production of higher concentrations of soil CO_2 and organic acids. These increase rates of carbonation, solution and hydrolysis. The highest rates of chemical weathering are associated with regions that have relatively high mean temperatures, and where high annual precipitation ensures the availability of water to act as a reactant and solvent. Topography affects temperature and water availability at regional and local scales, and the relationship between climate zones and weathering regimes is disrupted by mountain ranges. The mineral composition of the parent material determines the type of chemical weathering. Minerals such as olivine and pyroxene, which are rich in iron and other metallic ions, are very susceptible to breakdown by solution, oxidation and hydrolysis. Quartz, which contains few metallic ions, breaks down much more slowly. Physical properties of the parent material, such as jointing, bedding planes and fractures, increase the area available for chemical weathering and allows water to penetrate cracks and fissures.

MASS WASTING

Mass wasting refers to the various ways in which regolith, soil and rock move down-slope under the force of gravity. It does not include transport by moving water, air or ice, although these may act as lubricants, moving with the sediment in mudflows and avalanches, or causing periodic swelling and shrinkage of slope material. Mass wasting processes operate over a huge range of spatial and temporal scales. For example, individual particles may be transported a few centimetres over hundreds of years by soil creep, while landslides involving whole mountainsides may occur within a few minutes. The largest landslide on Earth is the Saidmarreh slide in south-west Iran. This took place 10,000 years ago and involved a mass of limestone 15 km long, 5 km wide and at least 300 m thick, which travelled several kilometres (Summerfield, 1991).

Forces acting on a block of slope material

One of the most obvious controls on the down-slope movement of slope material is the steepness of the slope. The gravitational force acts vertically downwards, but the slope material cannot move vertically downwards – it must either stay where it is or move in a down-slope direction. The forces acting on a block of weathered slope material are shown in Figure 4.2. The gravitational force can be divided into two components: a downslope component and a slope-normal component. The **down-slope component** exerts a **shear stress** on the block. This shear stress acts on the block in the same plane as the slope (parallel to the slope surface), exerting a down-slope pull on the block. The **slope-normal component** acts at right angles (normal) to the slope and resists movement by 'holding' the block in place. Another resisting force counteracting the down-slope component of gravity is frictional resistance between the block and the slope.

The relative size of the two components of the gravitational force is determined by slope angle. On gentle slopes, the slope-normal component is greater than the down-slope component. On steeper slopes the downslope component of gravity dominates, increasing the shear stress on the block, and promoting movement. Whether or not the block will actually move is dependent on the balance between the down-slope component (**driving force –** promoting movement) and the combined effects of the slope-normal component and frictional resistance (**resisting force –** resisting movement).

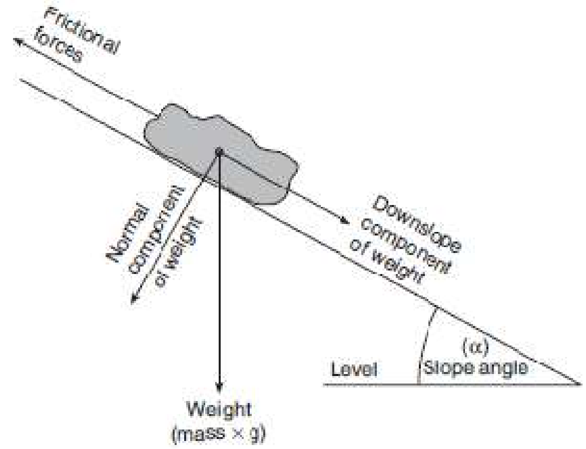


Figure 4.2 Driving and resisting forces acting on a block of slope material.

Shear stress and shear strength

Unfortunately, predicting the stability of slopes is not simply a matter of measuring slope angle. Things are greatly complicated by interactions between the particles forming the slope material, and the moisture held in the pore spaces between particles. This means that as well as causing down-slope movement, the applied shear stress can also cause material to deform. This leads to various types of flow, slumping and slow creep. Depending on the internal resistance of the slope material, and the friction with the slope itself, the application of a shear stress can result in deformation and/or sliding.

The resistance to deformation, or **shear strength**, depends on the properties of the material and the prevailing environmental conditions. Shear strength is partly determined by the amount of internal frictional resistance *within* the slope material. For example, shear strength increases when grains are tightly packed together. This increases the contact between individual grains and therefore the friction between them. **Cohesive forces** between particles can be very significant where silt and clay-size particles are involved. The physical and chemical structure of these particles leads to the development

of attractive **electrochemical forces** between them. Under some circumstances these can greatly increase the shear strength of slope materials. There are various dynamic processes that act as triggering mechanisms for mass movements. These include earthquakes, increased slope loading due to rainfall (Brooks and Richards, 1994), and wildfire, which alters the soil hydrological properties and reduces protection of the surface by vegetation (Prosser and Williams, 1998).

Other processes are also significant, such as slope loading caused by the accumulation of debris, changes in vegetation cover and slope steepening, as a result of undercutting at the base of the slope by fluvial erosion. Moisture content is very important on soil covered slopes because water can act as a lubricant when the pore spaces between particles become saturated. The pressure of the water – termed a **positive pore water pressure** – tends to force the particles apart. This reduces internal frictional resistance and therefore shear strength. By contrast, as the material dries out, only the smallest pores are filled with water and a **negative pore water pressure** develops. The water is held in these small pore spaces by strong suction forces. A negative pore water pressure can *increase* the shear strength of the material and is most effective for the smallest clay and silt-sized particles. Given the importance of pore water pressure, many landslides on soil-covered slopes are driven by rainstorms. Pore water pressure is highly dynamic, changing rapidly over time in response to inputs of precipitation.

Although the effectiveness of a given storm is partly dependent on its intensity and duration, a very significant factor is the moisture status of the slope material prior to the storm. Where there has been a high antecedent rainfall, it may only take a small event to trigger the down-slope movement of unstable materials.

Types of mass movement

Different types of mass movement are illustrated in Figure 4.3. Some of these, such as slides, falls and certain types of flow, occur over a short period of time (minutes to days) while non-rapid mass movements include heave, creep and solifluction. The term 'landslide' is in wide general use but actually describes several different types of mass movement. True **slides** are characterised by a well defined **shear plane** between the moving material and the slope. The shear plane is a two-dimensional surface which lies parallel to the slope for translational slides (Figure 4.3a) but has a concave form, a little like the upper surface of a spoon, for rotational slides (Figure 4.3b). Rotational slides tend to be more characteristic of thick layers of homogeneous, cohesive materials.

Flows differ from slides in that no shear plane exists. Although the greatest shear occurs at the base of the flow, nearly all the movement is within the body of the flowing mass. The liquid content of flows varies. While the moving material in mudflows is in an almost entirely fluid state, earthflows are part solid, part liquid, and move at almost imperceptible rates. Regolith **debris avalanches** are essentially dry flows, being distinguished from slides by the inter-particle movement within the flow. The more rapid types of flow usually have well defined boundaries. They frequently follow the line of pre-existing gullies, with a basin-shaped source area and a long, relatively narrow flow track that leads to a debris accumulation zone at the toe (Figure 4.3c). Slower earth flows create crescentic scars at the soil surface.

Falls (Figure 4.3d) involve the vertical displacement of loose boulders, rock fragments and finer material. **Solifluction** is the slowest type of flow and involves the down-slope movement of saturated slope materials. It is closely associated with **soil heave**, where slope material is vertically displaced, perhaps by the swelling and shrinking of clay particles, or the growth and melting of ice crystals. Repeated cycles lead to a net downward movement of soil material. Over longer time scales, regolith is subject to processes of **creep**, where material slowly moves down-slope under the force of gravity. Landslides and debris flows are the dominant sediment production process in many mountain drainage basins (Benda and Dunne, 1997) and represent an important source of bedload and suspended load to river channels. Rapid mass movements provide a spatially discrete input of coarse sediment to the channel, affecting channel morphology at a localised scale. These inputs also increase the supply of coarse sediment to the channel, leading to increased rates of transport and sediment storage in the channel network (Benda and Dunne, 1997).

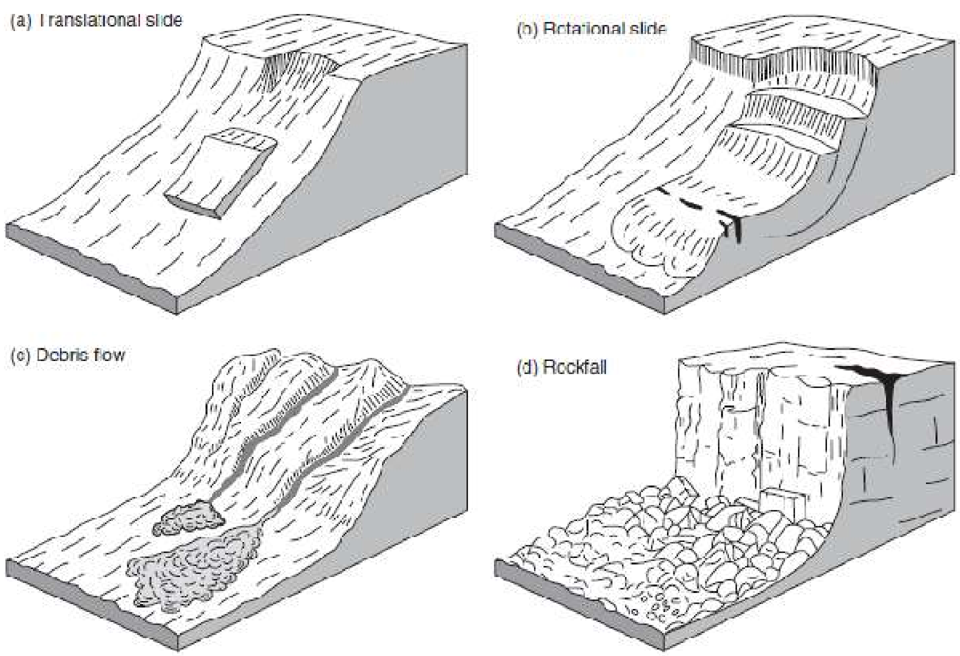


Figure 4.3 Examples of mass movement. (a) Translational slide. (b) Rotational slide or slump. (c) Debris flow. (d) Rockfall. After Huggett (2003).

WATER EROSION ON HILLSLOPES

Water erosion of the soil surface is brought about by the action of falling raindrops and surface flow, which may move as a sheet across the surface or be concentrated in rills or gullies. Subsurface flow is also significant in hillslope erosion. Soil erosion provides the main source of the fine suspended sediment that is transported by river channels (clays to fine sands). Where flow is concentrated, larger material can also be transported, for example where deep gullies erode into the coarser subsoil. There are many interrelated variables affecting rates of erosion, including climate, parent material, relief, tectonic setting, vegetation cover and human activity. For purposes of simplification, these can be considered in terms of the erosivity of the eroding agent and the erodibility of the soil surface (Figure

4.4). **Erosivity** is a measure of the capacity of an eroding agent, such as rainfall or overland flow, to erode the soil surface. It is dependent on the available kinetic energy (defined below), which is determined by factors such as rainfall intensity, raindrop size, flow depth and slope angle. **Erodibility** refers to the susceptibility of the soil surface to erosion and is dependent on the properties of the soil itself, such as soil texture (relative proportion of sand, silt and cohesive clay particles). As you will see later in this chapter, erodibility is also dependent on the amount and type of vegetative cover and on land use practices.

Rain splash erosion

Falling raindrops, in common with all moving objects, possess **kinetic energy**. This energy is often sufficient to detach soil particles when raindrops hit the soil surface, although a large proportion of the energy is used in compacting the surface and creating an impact crater. The amount of kinetic energy a moving object has is determined by the mass of the object (this relates to the size of the raindrop) and its velocity, as shown below: Kinetic energy □ where m □□mass and v □□velocity.

You have probably heard of the terminal velocity of an object. When a raindrop starts to fall it accelerates, due to the force of gravity, from an initial velocity of zero. This gravitational acceleration takes place at a rate of 9.81 m s–2, known as the gravitational constant, *g*. However, the raindrop does not continue to accelerate indefinitely, because it is also subject to a **drag force** as it moves through the atmosphere. This drag force acts vertically upwards, and increases with the velocity of the raindrop until it is equal and opposite to the gravitational force. This does not stop the raindrop, but it does stop it accelerating, with the result that the raindrop continues to fall at a constant, or terminal, velocity. Raindrops do not always reach their terminal velocity and factors such as wind speed and turbulence may increase or decrease their effectiveness when they land. Also important is the presence, percentage coverage and type of vegetation. This intercepts rainfall and breaks the fall of raindrops, reducing their kinetic energy before they reach the ground surface. Rain splash erosion is therefore most effective in areas where vegetation does not entirely cover the ground surface. When raindrops fall on a sloping surface, there is a net movement of material down-slope; this increases with slope angle.

Another important factor is the soil type – soils with a high silt/clay content offer the most resistance to erosion because of their cohesive nature. **Sheetwash erosion** When significant overland flow occurs, water flows over the surface in thin layers as so-called **sheet flow**. This is a somewhat misleading term because the flow is rarely of a uniform depth, being characterised by deeper, faster

threads of flow that result from micro-scale variations in the surface topography.

The down-slope flow of water exerts a shear stress on the soil surface. Erosion takes place when this shear stress is sufficient to overcome the resistance of the soil surface. Although the erosivity of the flow increases with depth and velocity, the shallow depth of overland flow and the roughness of the soil surface mean that the shear stress is not always sufficient to erode soil particles. As a result sheetwash is only really effective on steep slopes and smooth bare soil surfaces (Morgan, 2005). However, raindrop erosion is very effective as a detachment mechanism, allowing material to be entrained (set in motion). Since the transport of soil particles requires less energy than their initial entrainment, this material can be carried by the flow until it is deposited. The combined action of raindrop erosion and sheet erosion can therefore erode a significant volume of soil from large areas of sloping land. Sheetwash tends to be dominated by fine material with a diameter of less than 1 mm (Morgan, 2005), which contributes to the suspended load of rivers. Since soil is removed in thin layers, this type of erosion may go undetected for some time.

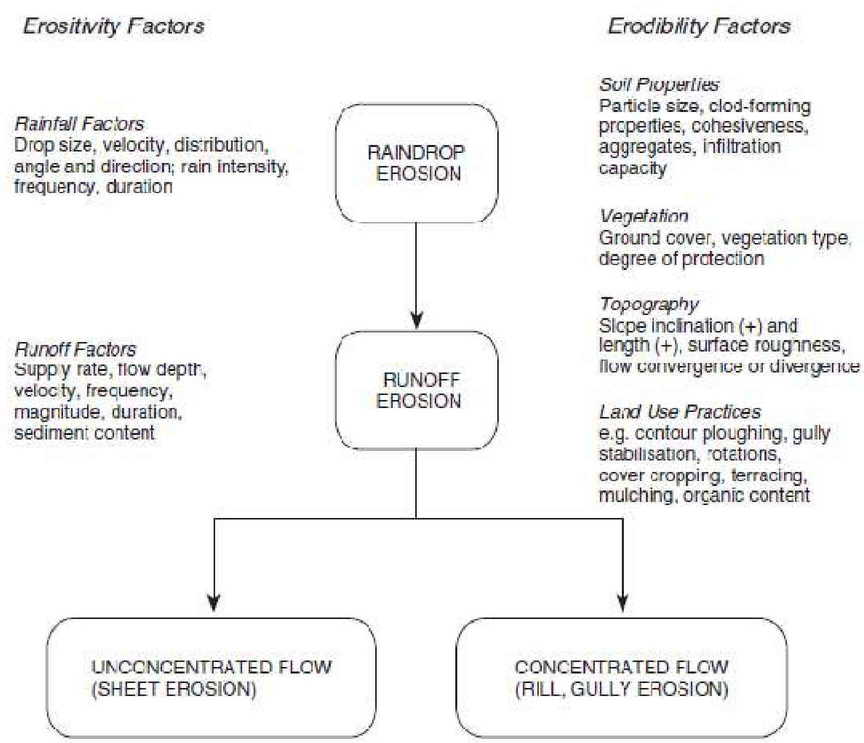


Figure 4.4 Summary of the principal factors controlling water erosion on hillslopes and sediment yield to river channels. After Cooke and Doornkamp (1990).

Rills

If the flow is sufficiently concentrated, a critical shear stress may be reached at which small micro-channels called **rills** start to form. Some well developed rills, formed in a road cutting, are shown in Plate 4.1. Rills vary in size with widths of between 50 mm and 300 mm and depths of up to 30 mm (Knighton, 1998). The critical conditions under which rills start to form can be considered in terms of a **critical shear stress** after Horton's (1945) theory of slope erosion by overland flow. It should be noted that this applies mainly to sparsely-vegetated dryland environments where intense rainfall and overland flow occur on a fairly regular basis. The diagram in Figure 4.5 represents overland flow occurring on a slope. The depth of flow increases with distance from the drainage divide, as flow accumulates in a down-slope direction (this has been exaggerated for clarity in the diagram). Since shear stress increases with depth, there must be a critical point on the slope at which the shear stress is great enough to allow incision to occur. This point is reached a **critical distance (Xc)** from the drainage divide, where the flow reaches a **critical depth (d)**. Xc varies from slope to slope according to the balance between erosivity and erodibility.

Above this point on the slope is a **belt of no erosion**. Incision can occur below this point, and parallel rills start to form in the **belt of active erosion**. Further up the slope these features tend to be discontinuous and ephemeral, being destroyed by inter-rill erosion or wall collapse. The eroded sediment is carried down-slope by the flow, reducing the energy available for further incision. If the transport capacity of the flow is exceeded, deposition starts to occur in the form of small fan-shaped features. Horton called this the **zone of deposition**. In the field, these zones are not as clearly defined as might appear from the diagram because soils are typically very heterogeneous. Even at the micro-scale, there is considerable variation from place to place in slope, roughness, infiltration capacity, cohesiveness and other factors affecting erodibility. As a result, complex spatial relationships exist between areas of erosion and areas of deposition.

Rills may develop into more permanent features under favourable conditions. They are significant in the initiation of new stream channels when network extension occurs, or where a surface has recently been exposed, perhaps as a result of glacial retreat or volcanic eruption. In order for a permanent channel to form, a sufficient concentration of flow is required. This can happen when one rill becomes dominant over neighbouring rills, incising at a faster rate, concentrating flow at depth and leading to further incision. Even when permanent channels do not form, rill erosion, together with sheetwash and rainsplash erosion between rills (inter-rill erosion), results in the net removal of material from slopes. The concentrated flow in rills can transport larger soil particles, and even small rock fragments (Poesen, 1987). Rills may account for much of sediment removal from a hillside, although this depends on the spacing of rills and the extent of the area affected (Morgan, 2005). This can lead to the loss of soil fertility and productivity when erosion proceeds more rapidly than soil formation. New soil is produced at a rate of a few millimetres a century, whereas a single storm can result in the removal of several centimetres of soil (Woodward and Foster, 1997).

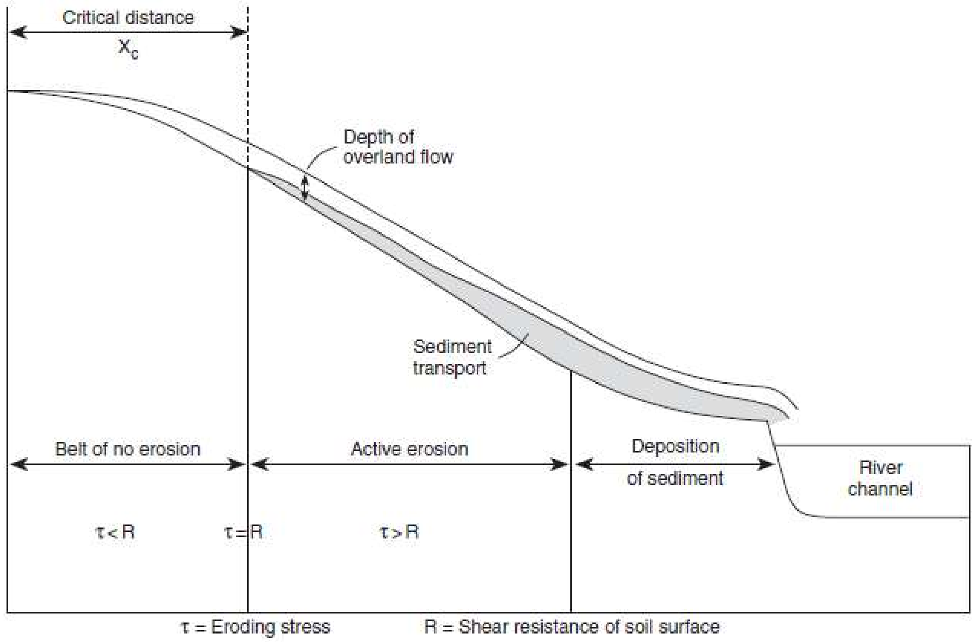


Figure 4.5 Horton's model of overland flow and rill formation. Adapted from Horton (1945).

Gullying

Gullies are relatively permanent ephemeral channels. They are most commonly found in arid and semi-arid environments, where precipitation is highly seasonal and vegetation cover is sparse. Gullies are morphologically different to stream channels, being relatively deep and narrow, with steep sidewalls and a stepped channel slope. They typically range in size from depths of 0.5 m up to 25 m or 30 m (Soil Science Society of America, 1996) although there is no clearly defined upper limit, and the distinction between large gullies and some ephemeral stream channels can be somewhat vague (Poesen *et al.*, 2002). Gullies are often connected to the river system and provide an effective link between upland areas and channels, allowing the rapid transmission of water and sediment into river systems. In dryland environments gully erosion is an important sediment source, contributing an average of 50-80 per cent of the overall sediment production (Poesen *et al*., 2002).

Smaller features, intermediate in size between rills and gullies, also exist. These are called **ephemeral gullies** and an example is shown in Plate 4.2. Ephemeral gullies are defined by the Soil Science Society of America (1996) as small channels that are eroded by concentrated overland flow and that can easily be filled by normal tillage, only to reform in the same location by additional runoff events. **Arroyos** are gully-like features that are cut into debris-choked valleys. Evidence from many arid and semi-arid areas in the south-western United States suggests that these features formed rapidly after the mid nineteenth century as a result of increasing soil erosion. Associated with gullies are **badlands**: high-relief areas that are intensively dissected by gullies and are useless for agriculture.

Badlands form on unconsolidated sediments, or poorly consolidated rocks in sparsely vegetated areas and may be initiated by gully erosion. They are associated with arid and semi-arid climates but can also form in more humid climates. The gully head Erosion is focussed at the **gully head**, the sharp break in slope at the upslope end of the gully. In the dramatic example shown in Colour Plate 15, overland flow is occurring over a large area (the green area is totally submerged), and 'waterfalls' can be seen where the flow plunges over multiple gully heads. Overland flow erodes the lip of the gully head as the water flows over it before falling into the plunge pool at its base, where deepening and undercutting of the headwall take place. This undermines the headwall and allows the gully head to retreat further up-slope. The steep sidewalls of the gully head are highly susceptible to various types of mass movement, especially when saturated during extreme events. Subsurface processes are also very significant in gully head retreat. Subsurface flow moving towards the gully head can weaken the walls, and the development of piping is common. The collapse of pipes further contributes to gully head retreat. Under certain circumstances, gullies can extend rapidly upslope and tributary gullies may also form. A considerable depth of material can be removed over a short period of time, even during a single runoff event. In extreme cases, rates of retreat can be tens of metres a year (Bull and Kirkby, 2002). As a result gullying can lead to serious management problems, which are discussed in the following section.

Variables that control rates of retreat include slope steepness and the drainage area up-slope from the gully. The location of the gully head changes according to the rate of sediment supplied from up-slope and the rate at which erosion takes place at the gully head. Gullies can become partly infilled if net deposition occurs. Gully formation It was originally thought that gullies developed from enlarged rills. However, the process is rather more complex and various theories have been developed on the basis of observations. Gullies are thought to form when a break in the vegetation cover allows erosional hollows to form, in which water accumulates. If sufficient flow concentration occurs, an incipient gully head, or **headcut**, forms. Erosion is focused at this point and, if overland flow occurs on a regular basis, material is eroded at the headcut to deepen and enlarge the hollow. This produces a more permanent feature, further concentrating flow and leading to more erosion as a result of positive feedback (Leopold *et al*., 1964). Once a gully channel has formed, retreat takes place by the mechanisms described above, as long as the rate of erosion exceeds the sediment supply. Gully development is also associated with landslides that leave deep steep-sided scars. These can be subsequently occupied by flowing water (Morgan, 2005). Not all gullies develop exclusively by surface erosion. Subsurface flow is also important, and there are a number of observations of gullies that have formed as a result of pipe collapse in various soil materials and under different climatic conditions (Morgan, 2005). Harvey (1982) suggested that the occurrence of piping and tunnelling is mainly controlled by the characteristics of the soil materials. At depth, differential porosity and soil strength allow for the development of preferential flow paths and promote pipe development. The presence of deep tension and desiccation cracks allow concentrated overland flow to penetrate the soil surface.

UNIT – 15: MECHANICS OF SEDIMENT DEPOSIT: CHANNEL COMPETENCE, CAPACITY AND EFFICIENCY

CONTROLS ON CHANNEL ADJUSTMENT AND FORM

Flow and sediment supply both fluctuate through time, meaning that continuous adjustment takes place through the erosion, reworking and deposition of sediment. The flow and sediment regimes are called **driving variables** because they drive these processes. Along a given reach, channel adjustment is constrained within certain boundaries that are imposed by local conditions. For example, a sand-bed river flowing across a wide floodplain is able to adjust its form much more readily than a bedrock channel confined within a narrow gorge. Energy availability is also important, and channel adjustments are often limited for rivers that flow over low gradients, especially where cohesive banks are protected by vegetation. These constraints are called **boundary conditions** and include valley confinement, channel substrate, valley slope and riparian vegetation. Figure 8.1 provides a schematic representation of the driving variables and boundary conditions that influence channel form.

A channel is said to be 'in regime' when its form fluctuates around an equilibrium condition over the time scale considered. Not all channels are in regime, and there are many examples of non-regime, or disequilibrium, channels. This may be because the channel is evolving in response to long

term changes in the flow or sediment regime, caused by a change in one of the external basin controls. Examples include incising or aggrading channels and those that are undergoing a change in channel pattern. Alternatively, some bedrock and dryland channels may exist in a permanent state of disequilibrium because it is only during flood flows that adjustments take place. In such cases, low flows have little or no influence on the overall channel form. Many empirical relationships have been developed to relate 'regime dimensions' (e.g. channel width or depth), to control variables (e.g. bankfull discharge). It is important to realise that these regime dimensions represent an average and are not applicable to all channel types or flow regimes.

The available stream power along a given reach is determined by the discharge and valley slope. At the sub-reach scale there are spatial variations in energy expenditure, which result from variations in channel shape and resistance to flow. These in turn influence patterns of erosion and deposition. For example, energy and erosion potential are concentrated where the channel narrows. Conversely, flow resistance is increased by obstructions to flow such as boulders, bedforms, bars or woody debris, which can lead to localised deposition. There is therefore two-way feedback between channel form and flow hydraulics – form influences flow and flow influences form.

This point is well illustrated by the work of Ashworth and Ferguson (1986) on a glacially fed braided river in Arctic Norway. An intensive monitoring programme was carried out to make detailed measurements of channel morphology, velocity and shear stress, bedload size and transport rate, and the size of bed material. Ashworth and Ferguson's diagram of the feedbacks between channel processes, morphological changes and sediments is shown in Figure 8.2. Starting at the top left of this diagram is the discharge of the river, which is unsteady (varies over time). The irregular form of the channel creates non-uniform flow conditions over the rough channel bed. As a result, complex spatial variations are seen in velocity, which also changes over time.

At any point in the channel, the bed shear stress is determined by the vertical velocity profile. Rates of bedload transport are determined by bed shear stress as well as the size and amount of bed material that is available for transport. As with velocity and shear stress distributions, rates of bedload transport are spatially variable, and also change with time. Bedload transport may maintain the existing channel shape, size and pattern. Alternatively, channel form can be modified as a result of scour, fill and possible lateral migration. The nature of such changes is spatially variable, and in turn feeds back to influence the velocity distribution within the channel. Bedload transport also governs the size distribution and structure of bed sediments through selective entrainment and transport. The character of the bed material determines the roughness of the channel, in turn affecting the velocity distribution in the channel.

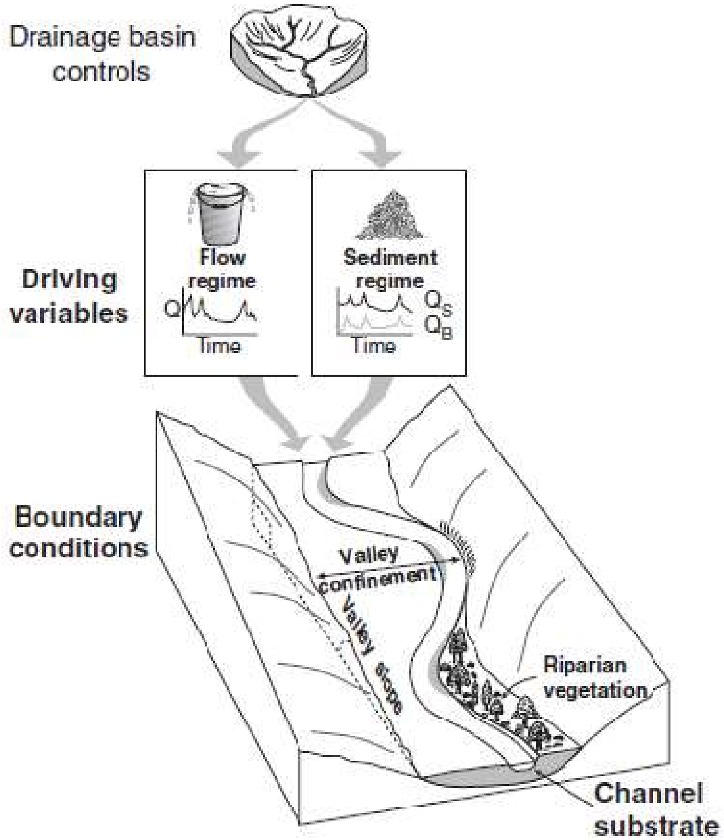


Figure 8.1 The driving variables and boundary conditions that control the form of a channel reach.

Driving variables

Flow regime The flow in natural river channels is unsteady, fluctuating through time in response to inputs of precipitation to the drainage basin. Characteristics of the flow regime include seasonal variations, flood frequency–magnitude relationships and the frequency and duration of low flows. Since discharge influences stream power, velocity and bed shear stress, the characteristics of the flow regime have an important influence on channel form.

Of morphological significance is the bankfull discharge. This was defined in and is the discharge at which the channel is completely filled with water. The bankfull discharge marks a morphological discontinuity between within-bank and out-of-bank flows. Since the flow in natural channels is unsteady, the bankfull discharge provides a representative flow. Channels are shaped by a range of flows. The geomorphological effectiveness of a given flood depends not only on its size, but also on the frequency with which it occurs. Large floods can carry out a considerable amount of geomorphological work. However, their comparative rarity means that the cumulative effect of smaller, more frequent flows may be more significant in shaping the channel. Box 3.3 provides a fuller explanation of the magnitude and frequency of channel-forming flows. Bankfull discharge (or the equivalent bankfull width) has often been used in developing statistical relationships between discharge and channel form parameters.

It is important to realise that bankfull discharge is actually quite difficult to define and that its frequency of occurrence varies considerably between different rivers.

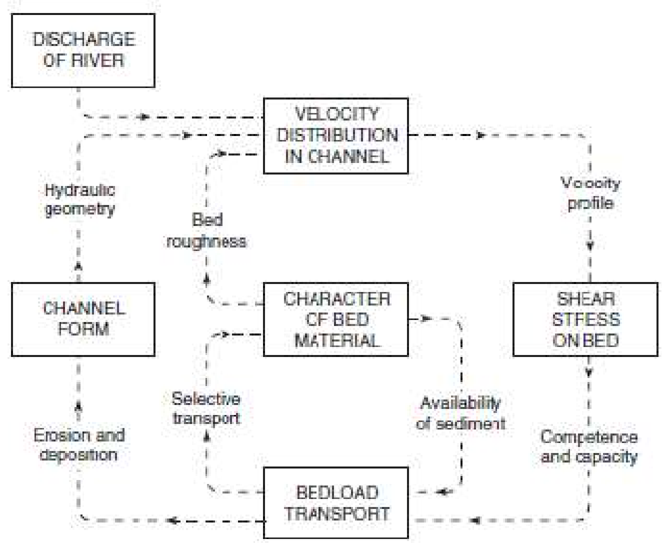


Figure 8.2 Interrelationships amongst form, flow and sediments in active gravel-bed rivers. After Ashworth and Ferguson (1986).

Sediment regime

The supply of sediment varies through time. It is not only the volume of sediment that is important but also its size distribution. As you will see later in this chapter, there are significant differences in the behaviour and morphology of bedload, suspended load and mixed load channels. Fluctuations in the volume and size of sediment are brought about by variations in sediment supply from the drainage basin and processes of sediment transfer through the channel network. As with the flow regime, it is the processes in the drainage basin, upstream from a given reach, that influence sediment supply.

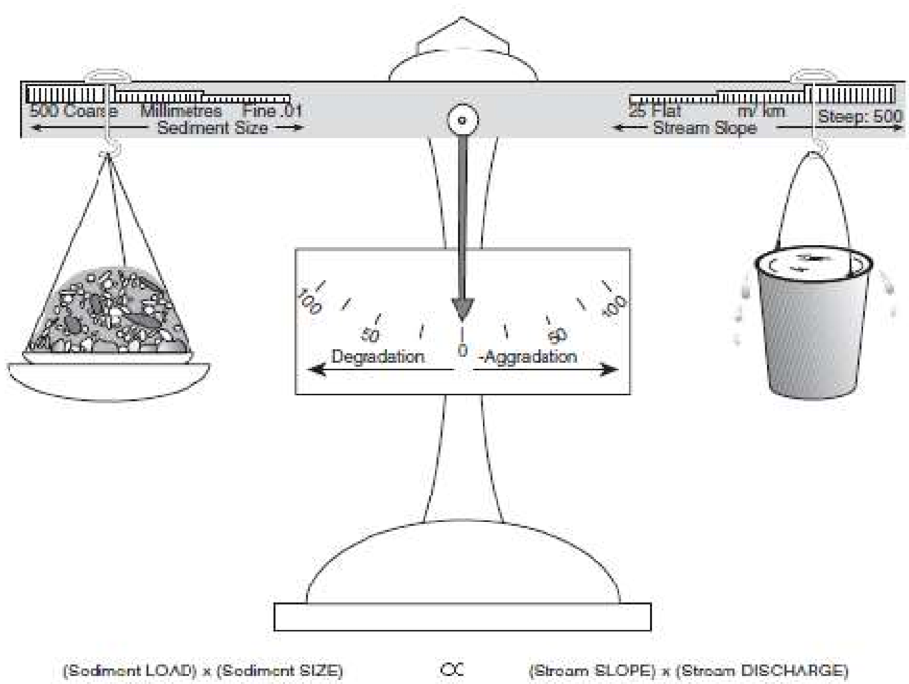


Figure 8.3 Illustration of the Lane balance between stream power (stream slope × discharge) and sediment supply (sediment load × sediment supply). This diagram, which originated as an unpublished drawing by W. Borland of the US Bureau of Reclamation, has been adapted from Brierley and Fryirs (2005).

The balance between stream power and sediment supply There is an important balance between the supply of bedload at the upstream end of a channel reach and the stream power available to transport it. This is known as the Lane balance, having first been described as a qualitative equation by Lane in 1955. This balance is illustrated schematically by the pair of scales shown in Figure 8.3. The left hand side of the scales represents the volume and size of sediment supplied to a channel reach over a given period of time. Balanced against this is the stream power available to transport it. This is determined both by the volume of water that enters the reach (over the same time period), and by the slope over which it flows. If the stream power is exactly sufficient to transport the sediment load, both sides of the scales are in balance and there is no net erosion or deposition along the reach. This is not to say that there is no erosion or deposition whatsoever, because these processes do occur at a localised scale in response to local variations in hydraulic conditions. Rather it means that, on balance, neither erosion nor deposition will predominate. An imbalance will occur if there is an increase in the volume or calibre of the sediment load in relation to the available stream power (sediment calibre is important

because it determines the flow competence required to transport it). This means that there is insufficient stream power to transport all the sediment, with the result that the excess is deposited along the reach. In this case, the balance tips towards **aggradation**, with net deposition occurring along the reach. Aggradation can be triggered in several ways, for example where the sediment supply is increased by upstream channel erosion, mass movement, or human activities such as mining. A particularly dramatic example of mining-induced aggradation along the Ok Tedi River in Papua New Guinea is shown in Colour Plate 16. Aggrading channels are characterised by numerous channel bars in a wide, shallow channel. Deposition within the channel may lead to the channel bed becoming elevated above the surface of the floodplain.

This, together with reduced channel capacity, increases the incidence of flooding and also promotes channel migration. A different situation arises when the stream power exceeds what is needed to transport the sediment load through the reach. This excess energy has to be expended somehow, so it is used to entrain sediment from the bed and erode the channel boundary. In this case **degradation** predominates. An example of a degrading channel is shown in Plate 8.1. Degradation can be caused by an increase in discharge, perhaps caused by an increase in flood frequency, or by a decrease in sediment supply. This can occur downstream from dams or where gravel mining has removed sediment from the river bed. The Lane balance is simplistic because much depends on the calibre of bed sediment within the reach. For example, no degradation can occur in a boulder-bed stream if the bed sediment is too coarse to be moved by the available stream power. This can be true even if the stream power exceeds the sediment supplied at the upstream end of the reach. Even when degradation does occur, another limitation of the equation is that it does not tell us *where* within the reach erosion will occur (Simon and Castro, 2003). This means that the equation cannot be used to predict the actual nature of channel change. For example, if the channel bed is more resistant to erosion than the banks, bank erosion is likely to be an initial adjustment. However, in a sand-bed channel with cohesive banks it is more likely that an initial adjustment would be scouring of the bed (Simon and Castro, 2003).

Resistance to erosion can be highly variable within a given reach, as can the specific stream power along that reach. This gives rise to spatially complex adjustments along the reach, even if there is net aggradation or net degradation along the reach as a whole.

Boundary conditions

Valley slope This refers to the downstream slope of the valley floor (as opposed to the slope of the channel itself) and determines the overall rate at which potential energy is expended along a given reach. The valley slope imposed on a given reach of channel is determined by a combination of factors including tectonics, geology, the location of the reach within the drainage basin and the long-term history of erosion and sedimentation along the valley. Although the overall energy available along a given reach is largely determined by the valley slope, it is possible for adjustments to occur that increase flow resistance at different scales (channel resistance, form resistance and boundary resistance). Different types of channel and floodplain morphology are associated with low, medium and high-energy environments.

Valley confinement

A channel may be defined as confined, partly confined, or unconfined, depending on how close the valley sides are. The degree of valley confinement is important for several reasons. In **confined** settings (Plate 8.2) channel adjustments are restricted by the valley walls, which also increase flow resistance. In addition, valley width influences the degree of slope–channel coupling that exists. Inputs of sediment from mass movements and other slope processes may exceed transport capacity, in turn influencing channel form. The episodic nature of mass movements means that these contributions can vary considerably over time.

In **partly confined**, some degree of lateral migration and floodplain development is possible. However, where the river comes against the valley wall or hillslope it is prevented from migrating further, which can lead to the development of over-deepened sections of channel. Stream power is also concentrated within the narrow valley and sections of the floodplain surface may be stripped during major floods. Where the hillslopes are a long way from the channel and have relatively little influence in contributing to the channel load, the channel is described as **unconfined**. Typically these settings are found in the lower reaches of rivers where there is very little interaction between channel and hillslopes.

Channel substrate

Considerable variations are seen in the form and behaviour of channels developed in different substrates. The substrate determines how resistant the channel is to the erosive force of the flow. It also influences boundary roughness, and therefore flow resistance. Alluvial channels formed in sand and gravel are generally more easily adjusted than those with cohesive silt and clay substrates. This is because the individual particles can be entrained at relatively low velocities, so non-cohesive substrates tend to be associated with wider, shallower cross-sections and faster rates of channel migration. Bedrock and mixed bedrock-alluvial channels are influenced over a range of scales by various geological controls.

Riparian vegetation

Vegetation on the banks and bed of river channels controls channel form in various ways. It often acts to protect and strengthen the banks, and research has shown that a dense network of roots can increase erosion resistance by more than a factor of ten. As a result, channels with vegetated banks are often narrower than those with non-vegetated banks under similar formative flows. This effect is most marked for densely vegetated banks (Hey and Thorne, 1986). Flow resistance can also be increased by vegetation growing on the bed and banks, as well as by woody debris (fallen trees and branches) that enters the channel from the banks.

An interesting example of the influence of riparian vegetation on channel form is provided by the Slesse Creek, British Columbia, Canada, and is reported by Millar (2000) and MacVicar (1999). The Slesse Creek drains an area of 170 km^2 within the Fraser River basin, flowing southwards from the United States into British Columbia. The reach shown in Plate 8.3 is approximately 2 km north of the US–Canadian border. This photograph was taken in 1940 and shows a stable meandering channel with a width of approximately 30 m and bordered by native forest vegetation. During the 1950s and 1960s logging of this stream bank vegetation took place along the British Columbia part of the channel. Subsequent to the removal of forest, the channel has developed an unstable braided form, widening to approximately 150 km. Millar (2000) suggests that these changes are mainly attributable to reduced bank stability, resulting from logging along river banks. This is because there has been no logging in the upstream (US) part of the drainage basin, which is a protected conservation area, so it is assumed that there has been little change in the flow or sediment regimes.

Downstream changes

Downstream changes in slope, discharge, valley confinement, sediment supply and particle size give rise to different balances between erosion and deposition along different parts of the profile. This leads to downstream changes in channel and floodplain morphology. In general terms, the cumulative supply of sediment increases downstream but the available energy decreases. Figure 8.4(a) shows how selected channel parameters change through the sediment production, transfer and deposition zones.

The discharge in most river channels increases in a downstream direction, as a progressivelylarger area is drained. In order to accommodate the growing volume of flow, channel dimensions (width and depth) typically increase downstream, and are often accompanied by a slight rise in velocity. The way in which these parameters change with increasing discharge can be described by the hydraulic geometry of the channel. Further information on hydraulic geometry relationships is provided in Box 8.1. Downstream reductions in bed material size reflect differences in the way in which coarse and fine sedimen are transferred along the channel. In contrast to the relatively localised transport of bedload particles, fine material, carried in suspension, is transported over much greater distances.

Observations show that there is a general decline in sediment size along the channel. The main causes of this downstream reduction are widely recognised as being abrasion and selective transport (Rice and Church, 1998). Abrasion refers to the reduction in size of individual particles by chipping, grinding and splitting. Physical and chemical weathering processes are also significant in the pre-weakening of individual particles. Selective transport refers to the longer travel distances associated with smaller grains, which are more mobile. The rate of reduction in sediment size varies considerablyand downstream *increases* are often observed at several locations. The downstream decrease in sediment size is often disrupted by inputs of coarser material. These include material from bank erosion, inputs from tributaries, and colluvial material. Material entering the main stream from tributaries is typically coarser than that in the main channel (Knighton, 1998). This causes a sudden increase in sediment size followed by a progressive fining further downstream. Complex patterns of downstream size reduction are seen where slope channel coupling is strong and non-alluvial supplies are dominant. These include contributions from hillslopes (e.g. mass movements), the erosion of bedrock outcrops and glacial material (Rice and Church, 1998).

The channel slope is typically steepest in the headwaters, becoming gentler in the lower reaches. The resulting long profile of many rivers is concave in shape, although the degree of concavity varies. Downstream increases in discharge, together with a decrease in bed material size, mean that the load can be transported over progressively shallow slopes. Exceptions to this are seen in arid and semi-arid regions, where downstream conveyance losses and high rates of evaporation lead toa downstream reduction in discharge. In this case a straight or convex profile may develop, since increasingly steep slopes are needed to compensate for the downstream reduction in flow.

Irregularities are often seen in the long profile, for example flatter sections are associated with lakes and reservoirs, and steeper sections at the site of waterfalls (see Figure 8.4b). In addition, there is often a change in the channel slope where tributaries join the main channel, because of the sudden increase in discharge. In tectonically active areas, where rates of uplift may be similar to erosion rates, rivers are in a state of dynamic equilibrium constantly trying to 'catch up' with tectonically driven changes.

CHANNEL ADJUSTMENT

Time scales of adjustment

Different components of a channel's morphology (e.g. bedforms, cross-sectional shape, slope) change over different time scales (Figure 8.5). This is because some components are more readily adjusted than others. For example, bedforms in a sand-bed channel are rapidly modified by a wide range of flows. Adjustments to channel width and depth take place over months to years, planform adjustments occur over tens to hundreds of years, while changes in the long profile may take thousands of years. Morphological adjustments therefore tend to lag behind the changes that cause them. This means that it can be difficult to link processes of flow and sediment transport with channel dimensions and form.

Channel form is directly controlled by flow regime and sediment supply. This chapter will deal with relatively short-term adjustments that are directly influenced by the flow and sediment regimes. These all act as controls on the flow and sediment regimes and, through a complex sequence of adjustments, lead to long-term changes in channel form and behaviour.

How adjustments are made

Channel form and behaviour reflect the driving variables and boundary conditions influencing a givenchannel reach. These controls also influence the ways in which channel adjustments are made. There are potentially four **degrees of freedom**, or variables, that can be modified: channel cross-section, slope, planform and bed roughness. Modifications to the cross-sectional size and shape are associated with changes in width and depth of the channel by processes such as bank erosion, incision of the bed, or aggradation. Channel slope can be adjusted in different ways. Negative feedback reduces the slope of steeper sections by erosion, and the slope of flatter sections is increased by deposition. Increases or decreases in channel length also affect channel slope, as illustrated in Figure 8.6. There are several different types of channel planform adjustments. These include lateral migration, meander bend development, reworking of bars, and even wholesale shifts of the channel to a new course. Finally, changes in bed roughness are brought about when the channel rearranges bed material, for example, in sand-bed channels, where bedforms are modified in response to changes in flow conditions. Mutual interrelationships exist between these variables, with adjustments made to one affecting one or more of the others. For instance, the formation of a meander cut-off alters the channel planform as well as increasing channel slope.

The influence of the driving variables and boundary conditions often reduces the degrees of freedom that a particular channel has. In the case of the mixed bedrock-alluvial channel shown in Colour Plate 3, depth increases are greatly restricted by the rock bed of the channel. On the other hand, the alluvial banks allow the channel to be widened much more easily. However, reductions in cross-sectional size by deposition may be limited if the channel has degradational tendencies.

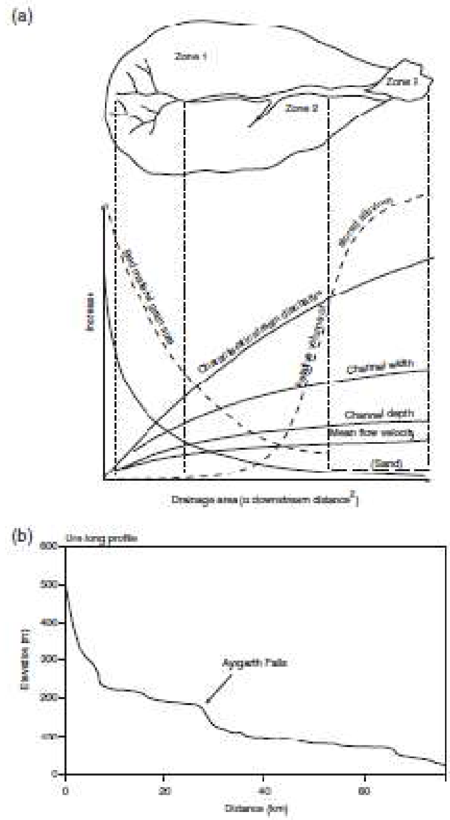


Figure 8.4 (a) Schematic representation of the variation in channel properties through a drainage basin. After Church (1992) and based on a concept of Schumm (1977). (b) Long profile of the River Ure, North Yorkshire, England. The Aysgarth Falls are a closely spaced series of three low waterfalls. After Coulthard *et al.* (2005).

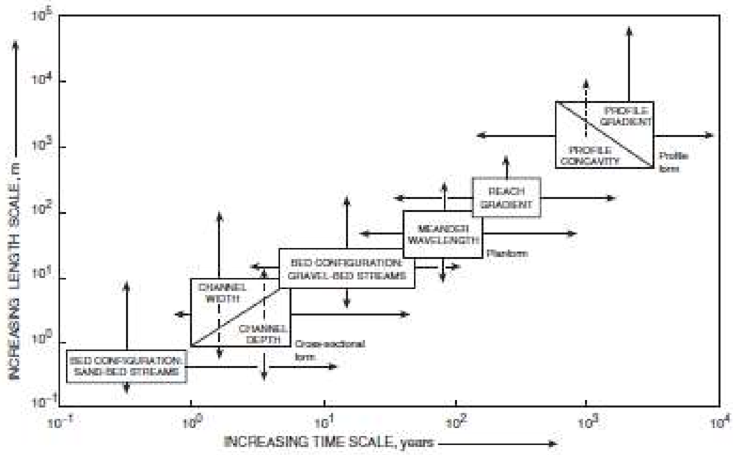


Figure 8.5 Schematic diagram of the time scales of adjustment of various channel form components with given length dimensions in a hypothetical basin of intermediate size. After Knighton (1998).

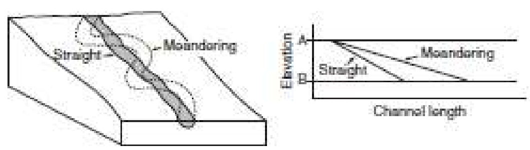


Figure 8.6 The development of meanders between two points, A and B, increases channel length. This reduces the channel slope between A and B.

UNIT – 16: SEDIMENT DEPOSITION- PROCESSES, NATURE AND CHARACTERISTICS OF FLUVIAL DEPOSITS; FLOOD PLAIN AND DELTAIC PLAIN DEPOSITS

CHANNEL GEOMORPHIC UNITS

Geomorphic units are features that form at the subchannel scale and can be erosional or depositional in origin. For instance, braided channels contain numerous mid-channel bars, while bedrock channels are associated mainly with erosional features such as potholes and bedrock steps, although bars can also form if sufficient bed sediment is available. Geomorphic units also affect hydraulic processes, and provide a range of different habitats for in-stream flora and fauna.

Bars

Bars are in-channel accumulations of sediment which may be formed from boulders, gravel, sand or silt. Bars can be divided into two broad groups: **unit bars** and **compound bars** (Smith, 1974). Unit bars are relatively simple bar forms whose morphology is mainly determined by processes of deposition (Ashmore, 1991). The evolution of these simple bar forms into more complex forms is described by Smith (1974), who made observations of the Kicking Horse River, British Columbia, Canada. Compound bars have a more complex history, having been shaped by many episodes of erosion and deposition. When erosion occurs, the basic shape of the bar is trimmed and dissected. Church and Jones (1982) recognise four main types of unit bars. These are illustrated in Figure 8.7.

Longitudinal bars are elongated in the direction of flow. They form in the centre of the channel, typically where the channel is relatively wide. Bar growth is brought about by the accumulation of finer material, both in an upwards and in a downstream direction (Church and Jones, 1982). Longitudinal bars tend to taper off in a downstream direction (Robert, 2003).

Transverse bars are lobe shaped (lobate) with relatively steep downstream faces. They are commonly found where there is an abrupt channel expansion, and downstream from confluences (Church and Jones, 1982). Transverse unit bars are not usually attached to the banks (Robert, 2003). The channel junction bars shown in Figure 8.7 are transverse bars that are associated with the flow separation that occurs at channel confluences.

Point bars are a feature of most meandering channels and form on the inside of meander bends as a result of the secondary flow patterns that are associated with flow in curved channels. Point bars are elongated in the direction of flow, with a steep outer face. **Diagonal bars** are common in gravel-bed channels (Robert, 2003). These are bank-attached features that run obliquely across the channel. Diagonal bars may have a steep downstream front. Both longitudinal and transverse bars are closely related to mid-channel bars. The compound mid-channel bars that characterise braided channels often have a complex history (see Colour Plates 6 and 7 for examples of these compound bars). Two terms that are commonly used to describe complex bar forms are medial(or lingoid) bars and lateral bars (Robert, 2003).**Medial bars** are symmetrical, detached from the bank sand have a characteristic lobate shape. **Lateral bars** are attached to one bank and have an asymmetric shape. Both types of compound bars have complex evolutionary histories. **Boulder bars** form in channels that are dominated by coarse bedload. As you will see later in this chapter, different morphologies are associated with the islandst hat are associated with anabranching channels. These include sand ridges, excavated islands, bedrock bars and vegetated bars with a bedrock core.

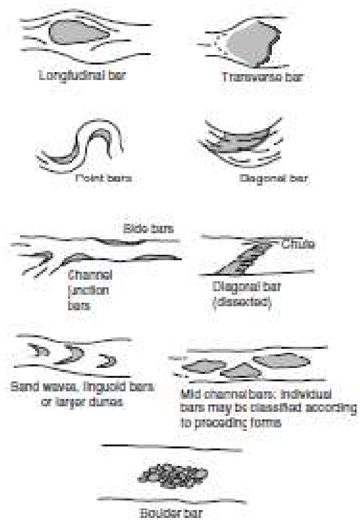


Figure 8.7 Types of channel bar. Adapted from Morisawa (1985), Church and Jones (1982) and Church (1992).

Benches

Benches are flat-topped, elongated, depositional features that form along one or both banks of channels. They are typically found on the inside of bends and along straight reaches, and are intermediate in height between the level of the channel bed and floodplain (Figure 8.8a). In bedrock and boulder-bed channels a boulder berm (bench composed of boulders) may form at the edge of the channel Benches can also form where flow separation occurs at the outer (concave) bank of tightly curving meander bends. This results in deposition and is illustrated in Figure 8.8(b). Erskine and Livingstone (1999) have observed sequences of adjacent benches along a bedrock-confined channel in the Hunter Valley, New South Wales, Australia. Rivers in this region have a very high flow variability, and each bench is associated with a different flow frequency. These benches are often eroded by catastrophic floods but are subsequently reconstructed by lower magnitude floods.

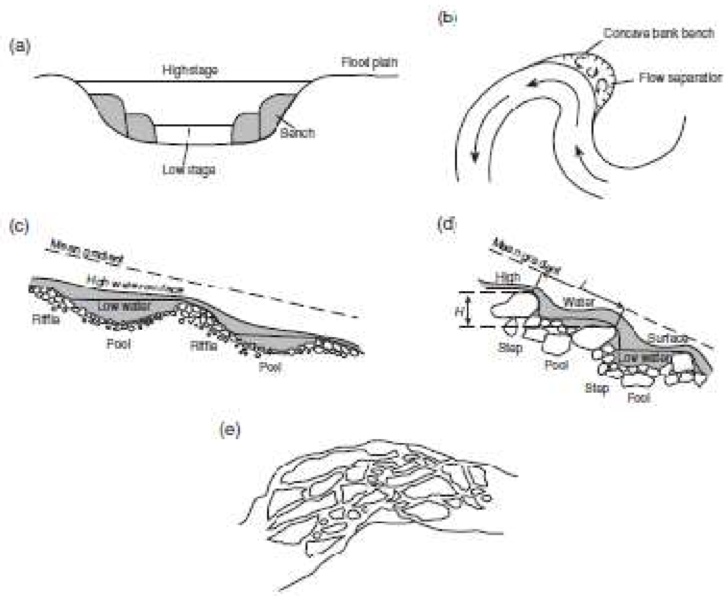


Figure 8.8 (a) Cross-sectional view of channel benches. (b) Concave bank bench. (c) Riffle–pool sequence. After Knighton (1998). (d) Step–pool sequence. After Knighton (1998). (e) Bedrock bars.

Riffle–pool sequences

The terms riffle and pool come from trout angling and refer to large-scale undulations in the bed topography. They are commonly found in gravel-bed channels with low to moderate channel slopes but do not tend to form in sand- or silt-bed channels (Knighton, 1998). The difference between riffles and pools is most obvious at low stages, when the flow moves rapidly over coarse sediment in the relatively steep riffle sections and more slowly through the deeper pools (Plate 8.5). The spacing from pool to pool, or riffle to riffle, is related to the width of the channel (and hence flow discharge). In most cases this is between five and seven times the channel width (Keller and Melhorn, 1978). A longitudinal section through a riffle–pool sequence is shown in Figure 8.8(c). This illustrates the differences in bed slope, bed material size and the slope of the water surface at high and low flows. At higher flows, the differences between riffles and pools are less obvious, with less variation in the water surface slope.

Riffle–pool sequences are found in straight, meandering and braided reaches. Analogous features are sometimes seen in ephemeral channels as regularly spaced accumulations of relatively coarse sediment, although there is little variation in the bed topography (Leopold *et al.*, 1966). In ecological terms, both riffles and pools provide important habitats. For example, certain species of fish lay their eggs in the spaces between the coarse gravels in riffles, while pools provide shelter and a suitable habitat for rearing young.

Various theories have been put forward to explain how riffle–pool sequences are maintained. Keller (1972) introduced a theory of velocity reversal. This suggests that the flow velocity increases at a faster rate in pool sections than in riffles as the discharge approaches bankfull. The higher shear stresses that develop in the pools lead to scouring of coarse material, which is deposited immediately downstream to form riffles. However, there is conflicting evidence to support this theory. Several researchers have shown that pools have a larger cross-sectional area of flow than riffles during most flow conditions. In order to ensure continuity of flow, pools should therefore have lower cross-sectional velocities. For example, Carling (1991) made observations on the River Severn, England. These indicated that neither the cross-sectional average velocity nor the near-bed shear velocity were noticeably greater in pools than riffles during overbank/near overbank conditions. Instead, there was a tendency for average hydraulic variables in riffles and pools to become more similar as the discharge increased.

Other theories have also been put forward. For example, field and laboratory measurements have shown that riffle surfaces tend to experience more turbulent flows. As a result, a tightly packed and interlocked bed surface develops at riffles. This is brought about by the vibration of particles and occasional particle transport during relatively low flows. In contrast, pools experience less near-bed turbulence during low flows and do not develop the same type of resistant bed structure (Robert, 2003). This means that critical bed shear stresses for sediment entrainment are higher in riffles than in pools. The riffles therefore tend to be maintained as topographic high points, while scouring occurs at pools (Robert, 2003).

Steps and pools

Steps and pools (Figure 8.8d and Plate 8.6) often characterise steep, upland channels and have been observed in a wide range of humid and arid environments. The steps are formed from coarser material and form vertical drops over which the flow plunges into the deeper, comparatively still water of the pool immediately downstream. Steps are relatively permanent features and consist of a framework of larger particles that is tightly packed with finer material. In forested catchments, woody debris has been observed to form part of the structure of steps. Steps and pools can also form in bedrock channels. Like riffles and pools, step–pool sequences are most apparent during low-flow conditions as they tend to be drowned out at higher flows. It is also during low-flow conditions that step–pool systems offer the most flow resistance. There is a considerable dissipation of energy as flow cascades over each step and enters the relatively still pools (Bathurst, 1993).

The spacing of steps and pools has been widely reported as being, on average, two to three times the channel width. Pools also tend to become more closely spaced as the slope increases. The height of steps appears to increase with the size of the bedload (Chin, 1999). Channels in which step–pool sequences form typically have a wide range of sediment sizes, from fine gravel to large boulders. Laboratory-based simulations indicate that step–pool sequences probably form during large floods, which mobilise the coarsest sediment. One theory suggests that, when the coarsest 'keystones' come to rest, they act as a barrier, leading to the accumulation of finer sediment. Downstream from this, the flow of water over the step scours a pool (Knighton, 1998).

Rapids and cascades

Like step–pool sequences, these are associated with steep channel gradients. Rapids are characterised by transverse, rib-like arrangements of coarse particles that stretch across the channel, while cascades have a more disorganised, 'random' structure. Rapids and cascades are stable during most flows because only the highest flows are competent to move the coarser cobbles and boulders that form the main structure.

Potholes

These deep, circular scour features are formed in bedrock channel reaches by abrasion.

Bedrock bars

In incised bedrock channels, the flow sometimes moves around bedrock bars. These form when multiple sub-channels are incised into the bedrock substrate, leaving 'islands' or bedrock bars between them. Bedrock bars may form the core of a bedrock-alluvial bar, which becomes covered by a layer of sediment on which vegetation becomes established.

FLOODPLAIN MORPHOLOGY

Processes of floodplain formation

The morphology of floodplains is intimately linked with the form and behaviour of the river channels that shape them. Various processes of deposition, reworking and erosion are involved in the formation and development of floodplains. Sediment accumulates on floodplain surfaces by various processes of accretion, the main ones being vertical, lateral and braid bar accretion (Nanson and Croke, 1992). **Lateral accretion** deposits are laid down by migrating rivers, which erode into the floodplain and lay down sediment in their wake. The accretion of point bar deposits can sometimes be seen as a series of concentric ridges on the inside of bends called **meander scrolls**. **Braid bar accretion** occurs when bars are abandoned and gradually become incorporated into the floodplain deposits. There are various ways in which this can happen, for example when a large flood lays down extensive bar deposits. Alternatively, bars may become abandoned when the main braid channels shift to another part of the valley. **Vertical accretion** deposits are composed of fine material that settles out of suspension when overbank flows inundate the floodplain. The increased area of contact, coupled with the roughness of the floodplain surface, greatly reduces flow velocities, and a thin layer of sediment is draped across the floodplain.

This displays a fining-upwards sequence, where the coarser particles, which settle out first, are overlain by progressively finer material. There is also a fining of sediment away from the channel, since only the very smallest particles are carried to the edge of the inundated area. Over a number of years the cumulative effect of overbank flows leads to the development of a vertical sequence of thin layers. Other, more localised, types of accretion can also be identified. For example, **counterpoint accretion** is associated with the deposition of concave bank benches at confined meander bends (see section on channel geomorphic units above). As an over-tightened meander bend migrates, bench deposits become incorporated into the floodplain.

Erosional processes include **floodplain stripping**, where entire sections of the floodplain surface are removed by high-magnitude flood events. Floodplain stripping is most likely to occur in relatively confined valley settings, where floodplain flows are concentrated between the valley walls. Other erosional processes include the formation of flood channels, which carry water during overbank flows. **Avulsion** involves a shift in the position of a channel and is a common process in braided reaches where the flow frequently abandons and reoccupies sub-channels. Avulsion can also involve the diversion of flow into a newly eroded channel cut into the floodplain. This type of avulsion is important in the development of anabranching channels. The morphology and development of floodplains is controlled by the driving variables and boundary conditions.

An important balance exists between the shear stress exerted by the flow and the resistance of the floodplain to erosion. Shear stress is closely related to specific stream power, and therefore to such controls as flow regime, valley slope and valley confinement. On the other side of the balance, resistance to erosion is largely determined by the cohesiveness of the floodplain sediments. An energy-based floodplain classification was proposed by Nanson and Croke (1992). This recognises three main classes of floodplain:

● ***High-energy non-cohesive floodplains*** are typically found in steep upland areas where the specific stream power in the channel at bankfull flow exceeds 300 W m–2. An example is shown in Plate 8.7. Lateral migration is often prevented by the coarseness of the floodplain sediment, which builds up vertically over time. These floodplains are disequilibrium features that are partly or completely eroded by infrequent extreme events.

● ***Medium-energy non-cohesive floodplains*** are formed from deposits ranging from gravels to fine sands. Specific stream power ranges from 10 to 300 W m–2. The main processes of floodplain construction are lateral point bar accretion (meandering channels) and braid bar accretion (braided channels). These floodplains are typically in dynamic equilibrium with the annual to decadal flow regime. Some of the features associated with medium-energy non-cohesive floodplains are illustrated in Figure 8.9(a).

● ***Low-energy cohesive floodplains*** are usually associated with laterally stable single thread or anastomosing channels. Formed from silt and clay, the dominant processes are vertical accretion of fine grained sediment and infrequent channel avulsions. Specific stream power at bankfull stage is generally less than 10 W m-2. Features associated with low energy cohesive floodplains are shown in Figure 8.9(b). These classes can be further subdivided, mainly on the basis of floodplain forming processes. They can also be related to the downstream reductions in stream power and sediment size discussed on pp. 124–128.

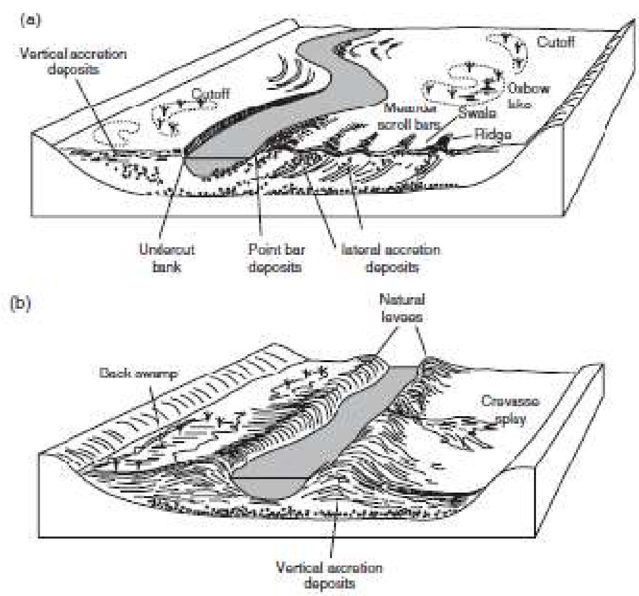


Figure 8.9 Some of the features associated with (a) medium-energy non-cohesive meandering floodplains, (b) low-energy cohesive floodplains. Adapted from Nanson and Croke (1992).

Floodplain geomorphic units

Levees

Levees are elongated, raised ridges that form at the channel–floodplain boundary during overbank flow events (Figure 8.9b). Moving across the boundary from channel to floodplain, there is a sudden loss of momentum because of the interaction between fast channel flow and slow floodplain flow. This results in the preferential deposition of material at the edges of the channel. Colour Plate 9 shows an anastomosing river in flood. Levees are clearly visible as the raised strips of land running along the channel margins. The height of levees is scaled to the size of the channel and their presence implies a relatively stable channel location (Brierley and Fryirs, 2005). These natural levees should not be confused with the artificial levees that are constructed along river banks for purposes of flood control.

Crevasse splays

Levees can be breached by floodwaters. This may lead to the formation of a crevasse splay, a fan-shaped lobe of sediment deposited when sediment-charged water escapes and flows down the levee (Figure 8.9a). If flow is sufficiently concentrated, a new channel may be cut and deepened by scour.

Backswamps. The build-up of sediment in the channel may mean that the channel is at a higher elevation than the surrounding floodplain. When levees are overtopped, water can enter the lower-lying area on the other side of the levee. This may be a depression or a swamp area characterised by wetland vegetation (Figure 8.9b). These are not exclusively associated with anabranching rivers and can also form at the valley margins of other channel types.

Flood channels

Flood channels are relatively straight channels that bypass the main channel. They have a lesser depth than the main channel and are dry for much of the time, only becoming filled with water as the flow approaches bankfull.

Floodouts

Floodouts are associated with dryland channels. They occur where floodwaters leave the main channel and branch out onto the floodplain in a number of distributor channels. This happens where low gradients, downstream transmission losses and high rates of evaporation lead to a downstream reduction in channel capacity. Channels may re-form downstream from the floodout if flow concentration is sufficient, forming a discontinuous channel. Alternatively the floodout may mark the channel terminus. Floodouts can also form where the channel is blocked by bedrock outcrops, fluvial, or Aeolian deposits such as sand dunes (Tooth, 1999).

Meander scroll bars

In some cases, former point bar deposits can be seen in the surface topography of the floodplain as **scroll bars**, with each scroll representing a former location of the point bar (Figure 8.9a). The undulating **ridge and swale topography** that results consists of higher ridges separated by topographic lows called swales. Migrating meanders do not always form scroll bars and the surface topography of these deposits may be relatively featureless.

Cut-offs

These are abandoned meander bends that have been short-circuited by the flow. Cut-offs become infilled over time by a process of abandoned channel accretion.

Palaeochannels

Palaeochannels are longer sections of abandoned channel. Like active channels, palaeochannels exhibit a wide range of different planforms. As time goes by, they gradually become infilled by abandoned channel accretion, the degree of infilling reflecting the age of the channel. The rate at which infilling occurs is dependent on factors such as the geometry of the

1.7 SELF-ASSESSMENT TEST

Discuss about mechanics of flow. Discuss about the factors affecting velocity.

Discuss about the mechanics of fluvial erosion. Discuss about the process of particle entrainment.

1.8 SUMMARIES AND KEY POINTS

- Mechanics of flow: uniform and non-uniform, steady and unsteady, laminar and turbulent, tranquil and rapid;
- Velocity and resistance: factors affecting velocity, flow resistance and viscosity;
- Stream power and energy;
- Mechanics of fluvial erosion: threshold of erosion, entrainment and bed erosion, headward erosion and channel lengthening, bank erosion by fluvial processes;
- River transportation: process of particle entrainment, initiation of sediment transport, selective theories of sediment transport, theories of equal mobility transport, transport of bed load and suspended load;
- Sediment load: nature and types of sediment load, sources of sediment load;
- Mechanics of sediment deposit: Channel competence, capacity and efficiency;
- Sediment deposition - processes, nature and characteristics of fluvial deposits; Flood plain and deltaic plain deposits

1.9 STUDY TIPS

Fundaments of fluvial morphometry – R. Charlton

Fluvial Processes in Geomorphology – Leopold, Wolman and Miller