Fundamentals of Physical Geography

Michael J. Pidwirny

Created by

Michael J. Pidwirny, Ph.D.,

Department of Geography, Okanagan University College
{PRIVATE}Email Suggestions : mpidwirny@okanagan.bc.ca

V e r s i o n - 1.31

http://www.geog.ouc.bc.ca/physgeog/home.html

Table of Contents

Preface

{PRIVATE**}Fundamentals of Physical Geography** is a work in progress. The pages that you can currently view are the result of several years of work. In the spring of 1996, I began creating a variety of web pages as learning supplements for two introductory Physical Geography courses I teach at Okanagan University College. From the fall of 1996 to spring of 1997, I repackaged these materials into online lecture notes for Distance Education versions of Geography 111 and 121. During 1997 and 1998, I made numerous revisions and additions to the online notes. I began working on the **Fundamentals of Physical Geography** online textbook in January of 1999. At the end of July 1999, the **Fundamentals of Physical Geography** website was officially put online. As of September 1st 2002, over 500,000 distinct computer clients (users) have accessed the website and more than 4.5 million page visits have occurred.

The **Fundamentals of Physical Geography** website is designed to be a **FREE** online textbook for University and College students studying introductory Physical Geography. **Version 1.3** of **Fundamentals of Physical Geography** contains over three hundred pages of information and more than four hundred 2-D illustrations, photographs, and animated graphics. Besides having the traditional text and 2-D graphics, this information source also has a number of animated graphics, an interactive glossary of terms, a study guide, web pages with links to other Internet resources related to Physical Geography, and a search engine to find information on the **Fundamentals of Physical Geography** website. The current purpose of this work is to **supplement** the printed textbooks used in Universities and Colleges with an information source that is interactive and rich in multimedia.

I would like to thank all those individuals who contacted me about typos and other errors in my webpages. A special thanks goes out to Allen Hutchinson for his diligence in providing me with numerous suggestions on how to improve the readability of this work.

Michael Pidwirny, September 8, 2002

Chapter 1: Introduction to Physical Geography

{PRIVATE}(a). Introduction to Geography

(b). Elements of Geography

(c). Scope of Physical Geography

(d). Geography as an Environmental Science

(e). History of Physical Geography

(f). Future of Physical Geography

(a) Introduction to Geography

{PRIVATE}The main objective of this online textbook is to introduce students to the field of knowledge known as Physical Geography. Physical Geography is a discipline that is part of a much larger area of understanding called Geography. Most individuals define Geography as a field of study that deals with maps. This definition is only partially correct. A better definition of Geography may be the study natural and human constructed phenomena relative to a spatial dimension.

The discipline of Geography has a history that stretches over many centuries. Over this time period, the study of Geography has evolved and developed into an important form of human scholarship. Examining the historical evolution of Ge ography as a discipline provides some important insights concerning its character and methodology. These insights are also helpful in gaining a better understanding of the nature of Physical Geography.

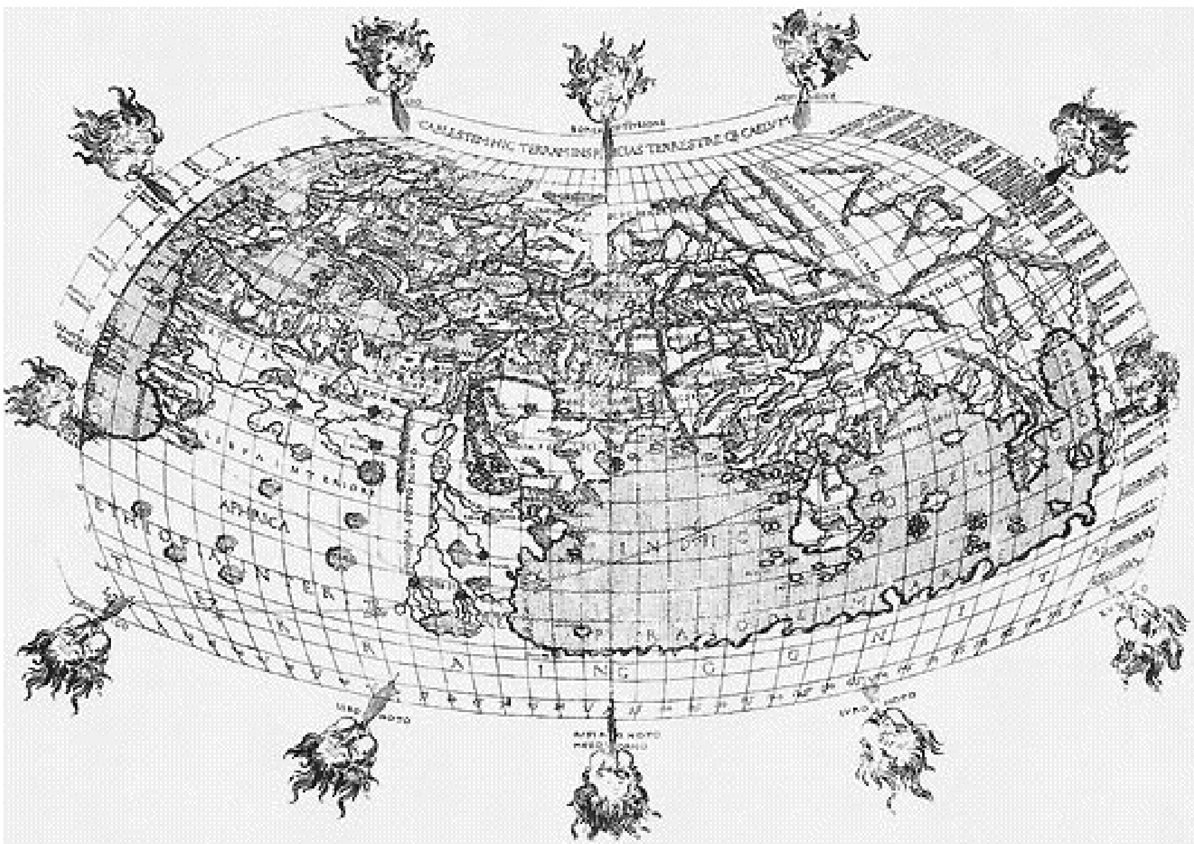
Some of the first truly geographical studies occurred more than four thousand years ago. The main purpose of these early investigations was to map features and places observed as explorers traveled to new lands. At this time, Chinese, Egyptian, and Phoenician civilizations were beginning to explore the places and spaces within and outside their homelands. The earliest evidence of such explorations comes from the archaeological discovery of a Babylonian clay tablet map that dates back to 2300 BC.

The early Greeks were the first civilization to practice a form of Geography that was more than mere map making or cartography. Greek philosophers and scientist were also interested in learning about spatial nature of human and physical features found on the Earth. One of the first Greek geographers was Herodotus (circa 484 - 425 BC). Herodotus wrote a number of volumes that described the Human and Physical Geography of the various regions of the Persian Empire.

The ancient Greeks were also interested in the form, size, and geometry of the Earth. Aristotle (circa 384 - 322 BC) hypothesized and scientifically demonstrated that the Earth had a spherical shape. Evidence for this idea came from observations of lunar eclipses. Lunar eclipses occur when the Earth casts its circular shadow on to the moon's surface. The first individual to accurately calculate the circumference of the Earth was the Greek geographer Eratosthenes (circa 276 - 194 BC). Eratosthenes calculated the equatorial circumference to be 40,233 kilometers using simple geometric relationships. This primitive calculation was unusually accurate. Measurements of the Earth using modern satellite technology have computed the circumference to be 40,072 kilometers.

Most of the Greek accomplishments in Geography were passed on to the Romans. Roman military commanders and administrators used this information to guide the expansion of their Empire. The Romans also made several important additions to geographical knowledge. Strabo (circa 64 BC - 20 AD) wrote a 17 volume series called "**Geographia**". Strabo claimed to have traveled widely and recorded what he had seen and experienced from a geographical perspective. In his series of books, Strabo describes the Cultural Geographies of the various societies of people found from Britain to as far east as India, and south to Ethiopia and as far north as Iceland. Strabo also suggested a definition of Geography that is quite complementary to the way many human geographers define their discipline today. This definition suggests that the aim of geography was to "**describe the known parts of the inhabited world**... **to write the assessment of the countries of the world [and] to treat the differences between countries**".

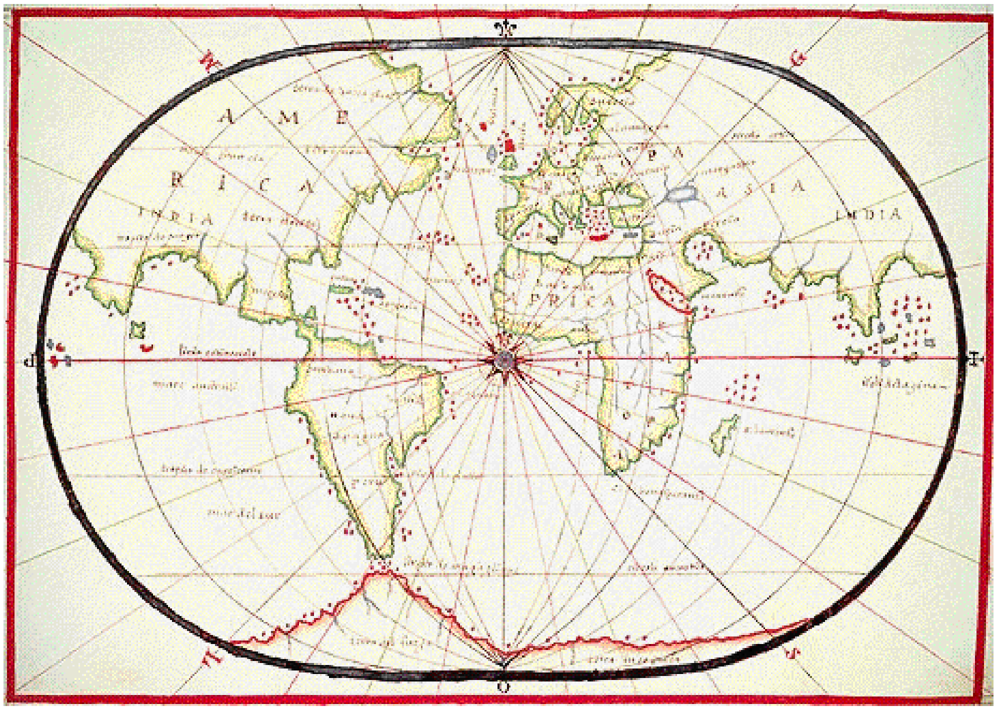
During the second century AD, Ptolemy (circa 100 - 178 AD) made a number of important contributions to Geography. Ptolemy's publication **Geographike hyphegesis** or "**Guide to Geography**" compiled and summarize much of the Greek and Roman geographic information accumulated at that time. Some of his other important contributions include the creation of three different methods for projecting the Earth's surface on a map, the calculation of coordinate locations for some eight thousand places on the Earth, and development of the concepts of geographical latitude and longitude (**Figure 1a-1**).



{PRIVATE}**Figure 1a-1:** This early map of the world was constructed using map making techniques developed by Ptolemy. Note that the map is organized with crisscrossing lines of latitude and longitude.

Little academic progress in Geography occurred after the Roman period. For the most part, the Middle Ages (5th to 13th centuries AD) were a time of intellectual stagnation. In Europe, the Vikings of Scandinavia were only group of people carrying out active exploration of new lands. In the Middle East, Arab academics began translating the works of Greek and Roman geographers starting in the 8th century and began exploring southwestern Asia and Africa. Some of the important intellectuals in Arab geography were Al-Idrisi, Ibn Battutah, and Ibn Khaldun. Al-Idrisi is best known for his skill at making maps and for his work of descriptive geography **Kitab nuzhat al-mushtaq fi ikhtiraq al-afaq** or "**The Pleasure Excursion of One Who Is Eager to Traverse the Regions of the World**". Ibn Battutah and Ibn Khaldun are well known for writing about their extensive travels of North Africa and the Middle East.

During the Renaissance (1400 to 1600 AD) numerous journeys of geographical exploration were commissioned by a variety of nation states in Europe. Most of these voyages were financed because of the potential commercial returns from resource exploitation. The voyages also provided an opportunity for scientific investigation and discovery. These voyages also added many significant contributions to geographic knowledge (**Figure 1a-2**). Important explorers of this period include Christopher Columbus, Vasco da Gama, Ferdinand Magellan, Jacques Cartier, Sir Martin Frobisher, Sir Francis Drake, John and Sebastian Cabot, and John Davis. Also during the Renaissance, Martin Behaim created a spherical globe depicting the Earth in its true three-dimensional form in 1492. Prior to Behaim's invention it was commonly believed in the Middle Ages that the Earth was flat. Behaim's globe probably influenced the beliefs of navigators and explorers of that time because it suggested that you could travel around the world.



{PRIVATE}**Figure 1a-2:** This map was constructed by Oliva in 1560. It describes the known world at this time and suggests that North America is part of Asia. Further exploration of the world would soon reject this idea.

In the 17th century, Bernhardus Varenius (1622-1650) published an important geographic reference titled **Geographia generalis** (**General Geography**: 1650). In this volume, Varenius used direct observations and primary measurements to present some new ideas concerning geographic knowledge. This work continued to be a standard geographic reference for about a 100 years. Varenius also suggested that the discipline of geography could be subdivided into three distinct branches. The first branch examines the form and dimensions of the Earth. The second sub-discipline deals with tides, climatic variations over time and space, and other variables that are influenced by the cyclical movements of the sun and moon. Together these two branches form the early beginning of what we collectively now call Physical Geography. The last branch of Geography examined distinct regions on the Earth using comparative cultural studies. Today, this area of knowledge is called Cultural Geography.

During the 18th century, the German philosopher Immanuel Kant (1724-1804) proposed that human knowledge could be organized in three different ways. One way of organizing knowledge was to classify its facts according to the type of objects studied. Accordingly, Zoology studies animals, botany examines plants, and geology involves the investigation of rocks. The second way one can study things is according to a temporal dimension. This field of knowledge is of course called History. The last method of organizing knowledge involves understanding facts relative to spatial relationships. This field of knowledge is commonly known as Geography.

Kant also divided Geography into a number of sub-disciplines. He recognized the following six branches: Physical, Mathematical, Moral, Political, Commercial, and Theological Geography.

Geographic knowledge saw strong growth in Europe and the United States in the 1800s. This period also saw the emergence of a number of societies interested in geographic issues. In Germany, Alexander von Humboldt, Carl Ritter, and Fredrich Ratzel made substantial contributions to human and physical geography. Humboldt's publication **Kosmos** (1844) examines the Geology and Physical Geography of the Earth. This work is considered by many academics to be a milestone contribution to geographic scholarship. Late in the 19th Century, Ratzel theorized that the distribution and culture of the Earth's various human populations was strongly influenced by the natural environment. The French geographer Paul Vidal de la Blanche opposed this revolutionary idea. Instead, he suggested that human beings were a dominant force shaping the form of the environment. The idea that humans were modifying the physical environment was also prevalent in the United States. In 1847, George Perkins Marsh gave an address to the Agricultural Society of Rutland County, Vermont. The subject of this speech was that human activity was having a destructive impact on land, especially through deforestation and land conversion. This speech also became the foundation for his book **Man and Nature** or **The Earth as Modified by Human Action**, first published in 1864. In this publication, Marsh warned of the ecological consequences of the continued development of the American frontier.

During the first 50 years of the 1900s, many academics in the field of Geography extended the various ideas presented in the previous century to studies of small regions all over the world. Most of these studies used descriptive field methods to test research questions. Starting in about 1950, geographic research experienced a shift in methodology. Geographers began adopting a more scientific approach that relied on quantitative techniques. The quantitative revolution was also associated with a change in the way in which geographers studied the Earth and its phenomena. Researchers now began investigating process rather than mere description of the event of interest. Today, the quantitative approach is becoming even more prevalent due to advances in computer and software technologies.

In 1964, William Pattison published an article in the **Journal of Geography** (1964, 63: 211-216) that suggested that modern Geography was now composed of the following four academic traditions:

Spatial Tradition - the investigation of the phenomena of Geography from a strictly spatial perspective.

Area Studies Tradition - the geographical study of an area on the Earth at either the local, regional, or global scale.

Human-Land Tradition - the geographical study of human interactions with the environment.

Earth Science Tradition - the study of natural phenomena from a spatial perspective. This tradition is best described as theoretical Physical Geography.

Today, the academic traditions described by Pattison are still dominant fields of geographical investigation. However, the frequency and magnitude of human mediated environmental problems has been on a steady increase since the publication of this notion. These increases are the result of a growing human population and the consequent increase in the consumption of natural resources. As a result, an increasing number of researchers in Geography are studying how humans modify the environment. A significant number of these projects also develop strategies to reduce the negative impact of human activities on nature. Some of the dominant themes in these studies include: environmental degradation of the hydrosphere, atmosphere, lithosphere, and biosphere; resource use issues; natural hazards; environmental impact assessment; and the effect of urbanization and land-use change on natural environments.

Considering all of the statements presented concerning the history and development of Geography, we are now ready to formulate a somewhat coherent definition. This definition suggests that Geography, in its simplest form, is the field of knowledge that is concerned with how phenomena are spatially organized. Physical Geography attempts to determine why natural phenomena have particular spatial patterns and orientation. This textbook will focus primarily on the **Earth Science Tradition**. Some of the information that is covered in this online textbook also deals with the alterations of the environment because of human interaction. These pieces of information belong in the **Human-Land Tradition** of Geography.

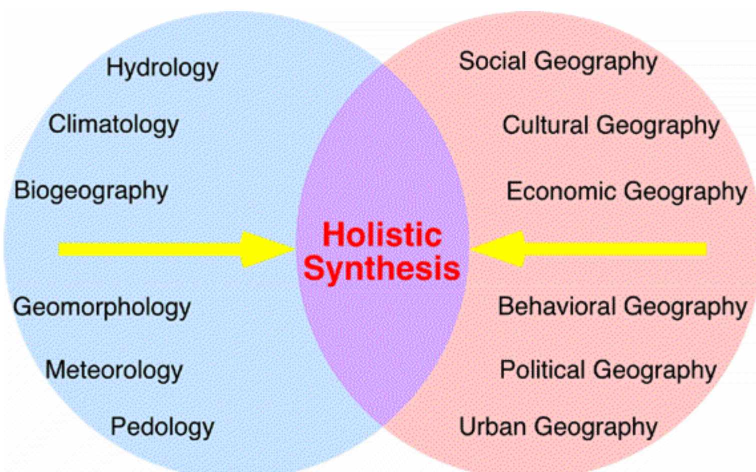
(b) Elements of Geography

{PRIVATE**}** In the previous section, we discovered that Geography consists of at least two different sub-fields of knowledge with similar methodology: Physical Geography and Human Geography. The following **table** also helps to make the differences between these two types of Geography more apparent. This table describes some of the phenomena or elements studied by each of these sub-fields of knowledge. Knowing what kinds of things are studied by Geographers provides us with a better understanding of the differences between Physical and Human Geography.

Table 1b-1: Some of the phenomena studied in Physical and Human Geography.

{PRIVATE**}Physical Geography**	**Human Geography**
Rocks and Minerals	Population
Landforms	Settlements
Soils	Economic Activities
Animals	Transportation
Plants	Recreational Activities
Water	Religion
Atmosphere	Political Systems
Rivers and Other Water Bodies	Social Traditions
Environment	Human Migration
Climate and Weather	Agricultural Systems
Oceans	Urban Systems

Geography is also discipline that integrates a wide variety of subject matter. Almost any area of human knowledge can be examined from a spatial perspective. **Figure 1b-1** describes some of the main subdisciplines within Human and Physical Geography. Physical Geography's primary subdisplines study the Earth's atmosphere (Meteorology and Climatology), animal and plant life (Biogeography), physical landscape (Geomorphology), soils (Pedology), and waters (Hydrology). Some of the dominant areas of study in Human Geography include: human society and culture (**Social and Cultural Geography**), behavior (**Behavioral Geography**), economics (**Economic Geography**), politics (**Political Geography**), and urban systems (**Urban Geography**).



{PRIVATE**}Figure 1b-1:** Major subdisciplines of Physical and Human Geography.

The graphic model in **Figure 1b-1** indicates that the study of Geography can also involve a holistic synthesis. Holistic synthesis connects knowledge from a variety of academic fields in both Human and Physical Geography. For example, the study of the enhancement of the Earth's greenhouse effect and the resulting global warming requires a multidisciplinary approach for complete understanding. The fields of Climatology and Meteorology are required to understand the physical effects of adding addition greenhouse gases to the atmosphere's radiation balance. The field of Economic Geography provides information on how various forms of human economic activity contribute to the emission of greenhouse gases through fossil fuel burning and land-use change. Combining the knowledge of both of these academic areas gives us a more comprehensive understanding of why this serious environmental problem occurs.

The holistic nature of Geography is both a strength and a weakness. Geography's strength comes from its ability to connect functional interrelationships that are not normally noticed in narrowly defined fields of knowledge. The most obvious weakness associated with the Geographical approach is related to the fact that holistic understanding is often too simple and misses important details of cause and effect.

(c) Scope of Physical Geography

{PRIVATE**}**We have now learned that Physical Geography examines and investigates natural phenomena spatially. In the previous section, we identified some of the key elements studied by Physical Geographers. Combining these two items, we can now suggest that **Physical Geography studies the spatial patterns of weather and climate, soils, vegetation, animals, water in all its forms, and landforms**. Physical Geography also examines the interrelationships of these phenomena to human activities. This sub-field of Geography is academically known as the **Human-Land Tradition**. This area of Geography has seen very keen interest and growth in the last few decades because of the acceleration of human induced environmental degradation. Thus, Physical Geography's scope is much broader than the simple spatial study of nature. It also involves the investigation of how humans are influencing nature.

Academics studying Physical Geography and other related Earth Sciences are rarely generalists. Most are in fact highly specialized in their fields of knowledge and tend to focus themselves in one of the following well defined areas of understanding in Physical Geography:

Geomorphology - studies the various landforms on the Earth's surface.

Pedology - is concerned with the study of soils.

Biogeography - is the science that investigates the spatial relationships of plants and animals.

Hydrology - is interested in the study of water in all its forms.

Meteorology - studies the circulation of the atmosphere over short time spans.

Climatology - studies the effects of weather on life and examines the circulation of the atmosphere over longer time spans.

The above fields of knowledge generally have a primary role in introductory textbooks dealing with Physical Geography. Introductory Physical Geography textbooks can also contain information from other related disciplines including:

Geology - studies the form of the Earth's surface and subsurface, and the processes that create and modify it.

Ecology - the scientific study of the interactions between organisms and their environment.

Oceanography - the science that examines the biology, chemistry, physics, and geology of oceans.

Cartography - the technique of making maps.

Astronomy - the science that examines celestial bodies and the cosmos.

(d) Geography as an Environmental Science

{PRIVATE**}**Webster's 9th Collegiate Dictionary defines Environment "... **as the complex of physical, chemical, and biotic factors (such as climate, soil, and living things) that act upon an organism or an ecological community and ultimately determines its form and survival**".

Both Human and Physical Geography provide an important intellectual background for studying the environment. Many environmental studies/science programs offered by Universities and Colleges around the world rely on the information found in various Geography courses to help educate their students about the state of the environment.

(e) History of Physical Geography

{PRIVATE**}**The nature of understanding in Physical Geography has changed over time. When investigating this change it becomes apparent that certain universal ideas or forces had very important ramifications to the academic study of Physical Geography. During the period from 1850 to 1950, there seems to be four main ideas that had a strong influenced on the discipline:

(1). Uniformitarianism - this theory rejected the idea that catastrophic forces were responsible for the current conditions on the Earth. It suggested instead that continuing uniformity of existing processes were responsible for the present and past conditions of this planet.

(2). Evolution - Charles Darwin's Origin of Species (1859) suggested that natural selection determined which individuals would pass on their genetic traits to future generations. As a result of this theory, evolutionary explanations for a variety of natural phenomena were postulated by scientists. The theories of **uniformitarianism** and **evolution** arose from a fundamental change in the way humans explained the universe and nature. During the 16th, 17th, and 18th centuries scholars began refuting belief or myth based exp lanations of the cosmos, and instead used science to help explain the mysteries of nature. Belief based explanations of the cosmos are made consistent with a larger framework of knowledge that focuses on some myth. However, theories based on science questioned the accuracy of these beliefs.

(3). Exploration and Survey- much of the world had not been explored before 1900. Thus, during this period all of the fields of Physical Geography were actively involved with basic data collection. This data collection included activities like determining the elevation of land surfaces, classification and description of landforms, the measurement of the volume of flow of rivers, measurement of various phenomena associated to weather and climate, and the classification of soils, organisms, biological communities and ecosystems.

(4). Conservation - beginning in the 1850s a concern for the environment began to develop as a result of the human development of once natural areas in the United States and Europe. One of the earliest statements of these ideas came from George Perkins Marsh (1864) in his book "**Man in Nature**" or "**Physical Geography as Modified by Human Action**". This book is often cited by scholars as the first significant academic contribution to conservation and environmentalism.

After 1950, the following two forces largely determined the nature of Physical Geography:

(1). The **Quantitative Revolution** - measurement became the central focus of research in Physical Geography. It was used primarily for hypothesis testing. With measurement came mapping, models, statistics, mathematics, and hypothesis testing. The quantitative revolution was also associated with a change in the way in which physical geographers studied the Earth and its phenomena. Researchers now began investigating process rather than mere description of the environment.

(2). The study of **Human/Land Relationships** - the influence of human activity on the environment was becoming very apparent after 1950. As a result, many researchers in Physical Geography began studying the influence of humans on the environment. Some of the dominant themes in these studies included: environmental degradation and resource use; natural hazards and impact assessment; and the effect of urbanization and land-use change on natural environments.

(f) Future of Physical Geography

{PRIVATE**}**The following list describes some of the important future trends in Physical Geography research:

(1). Continued development of Applied Physical Geography for the analysis and correction of human-induced environmental problems. A student of Applied Physical Geography uses theoretical information from the field of Physical Geography to manage and solve problems related to natural phenomena found in the real world.

(2). Remote Sensing - Advances in technology have caused the development of many new instruments for the monitoring of the Earth's resources and environment from airborne and space platforms (see three-dimensional image of hurricane Andrew, Landsat image of San Francisco Bay, Landsat image of Vancouver, British Columbia, and a space radar image of Victoria, British Columbia). The most familiar use of remote sensing technology is to monitor the Earth's weather for forecasting. Also see **section** 2d.

(3). Geographic Information Systems - A geographic information system (GIS) merges information in a computer database with spatial coordinates on a digital map. Geographic Information Systems are becoming increasingly more important for the management of resources. Also see **section** 2e.

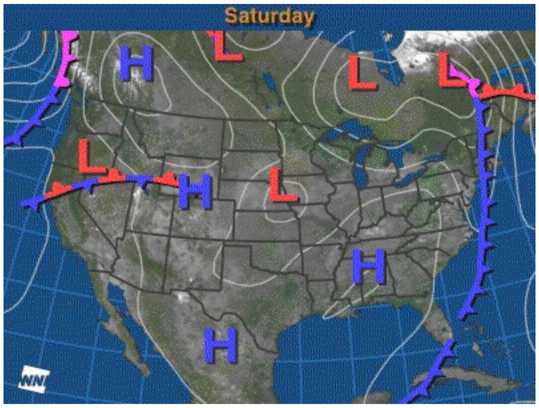
2) Maps, Remote Sensing, and GIS

(a) Introduction to Maps

{PRIVATE**}Introduction**

A map can be simply defined as a graphic representation of the real world. This representation is always an abstraction of reality. Because of the infinite nature of our Universe it is impossible to capture all of the complexity found in the real world. For example, topographic maps abstract the three-dimensional real world at a reduced scale on a two-dimensional plane of paper.

Maps are used to display both cultural and physical features of the environment. Standard topographic maps show a variety of information including roads, land-use classification, elevation, rivers and other water bodies, political boundaries, and the identification of houses and other types of buildings. Some maps are created with very specific goals in mind. **Figure 2a-1** displays a weather map showing the location of low and high pressure centers and fronts over most of North America. The intended purpose of this map is considerably more specialized than a topographic map.



{PRIVATE**}Figure 2a-1:** The following specialized weather map displays the surface location of pressure centers and fronts for Saturday, November 27, 1999 over a portion of North America.

The art of map construction is called cartography. People who work in this field of knowledge are called cartographers. The construction and use of maps has a long history. Some academics believe that the earliest maps date back to the fifth or sixth century BC. Even in these early maps, the main goal of this tool was to communicate information. Early maps were quite subjective in their presentation of spatial information. Maps became more objective with the dawn of Western science. The application of scientific method into cartography made maps more ordered and accurate. Today, the art of map making is quite a sophisticated science employing methods from cartography, engineering, computer science, mathematics, and psychology.

Cartographers classify maps into two broad categories: reference maps and thematic maps. Reference maps normally show natural and human-made objects from the geographical environment with an emphasis on location. Examples of general reference maps include maps found in atlases and topographic maps. Thematic maps are used to display the geographical distribution of one phenomenon or the spatial associations that occur between a number of phenomena.

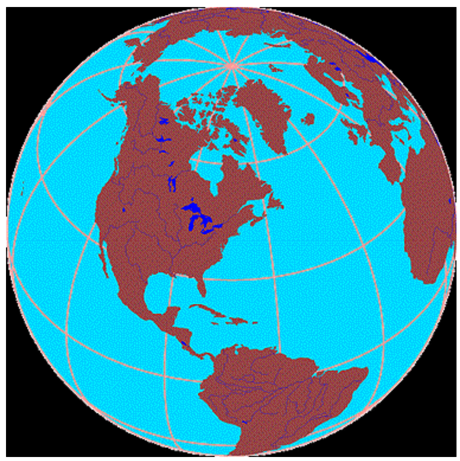
Map Projection

The shape of the Earth's surface can be described as an ellipsoid. An ellipsoid is a three-dimensional shape that departs slightly from a purely spherical form. The Earth takes this form because rotation causes the region near the equator to bulge outward to space. The angular motion caused by the Earth spinning on its axis also forces the polar regions on the globe to be somewhat flattened.

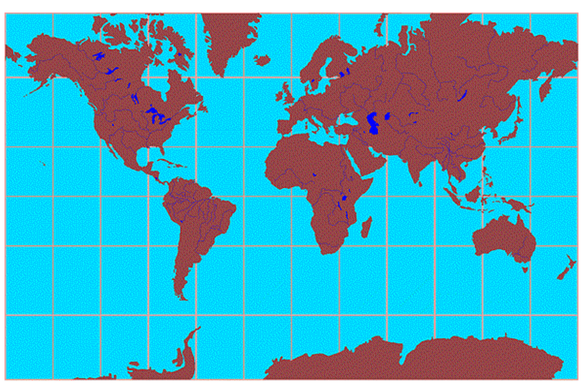
Representing the true shape of the Earth's surface on a map creates some problems, especially when this depiction is illustrated on a two-dimensional surface. To overcome these problems, cartographers have developed a number of standardized transformation processes for the creation of two-dimensional maps. All of these transformation processes create some type of distortion artifact. The nature of this distortion is related to how the transformation process modifies specific geographic properties of the map. Some of the geographic properties affected by projection distortion include: distance; area; straight line direction between points on the Earth; and the bearing of cardinal points from locations on our planet.

The illustrations below show some of the common map projections used today. The first two-dimensional projection shows the Earth's surface as viewed from space (**Figure 2a-2**). This orthographic projection distorts distance, shape, and the size of areas. Another serious limitation of this projection is that only a portion of the Earth's surface can be viewed at any one time.

The second illustration displays a Mercator projection of the Earth (**Figure 2a-3**). On a Mercator projection, the north-south scale increases from the equator at the same rate as the corresponding east-west scale. As a result of this feature, angles drawn on this type of map are correct. Distortion on a Mercator map increases at an increasing rate as one moves toward higher latitudes. Mercator maps are used in navigation because a line drawn between two points of the Earth has true direction. However, this line may not represent the shortest distance between these points.

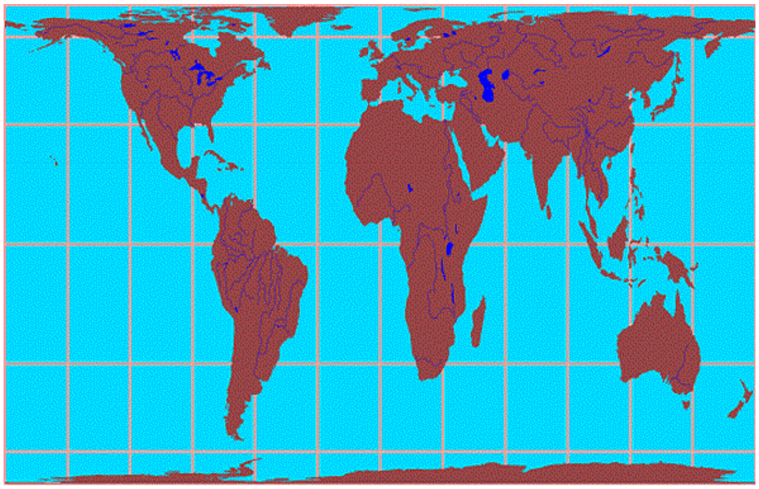


{PRIVATE}**Figure 2a-2:** Earth as observed from a vantage point in space. This orthographic projection of the Earth's surface creates a two-dimensional representation of a three-dimensional surface. The orthographic projection distorts distance, shape, and the size of areas.



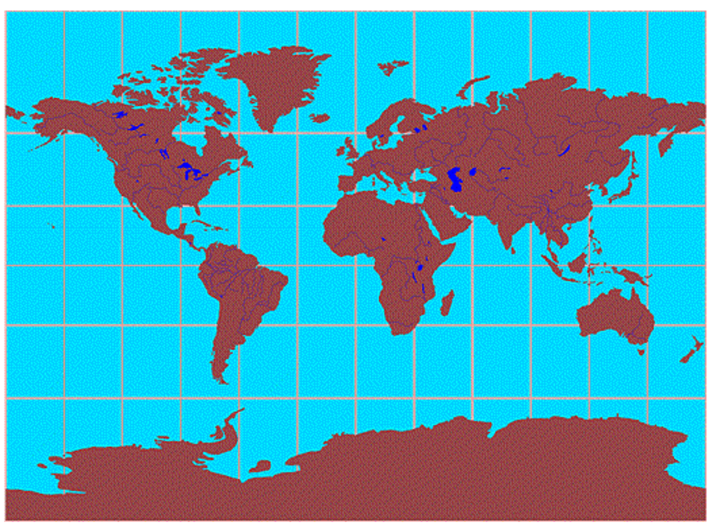
{PRIVATE}**Figure 2a-3:** Mercator map projection. The Mercator projection is one of the most common systems in use today. It was specifically designed for nautical navigation.

The Ga ll-Peters projection was developed to correct some of the distortion found in the Mercator system (**Figure 2a-4**). The Mercator projection causes area to be gradually distorted from the equator to the poles. This distortion makes middle and high latitude countries to be bigger than they are in reality. The Gall-Peters projection corrects this distortion making the area occupied by the world's nations more comparable.



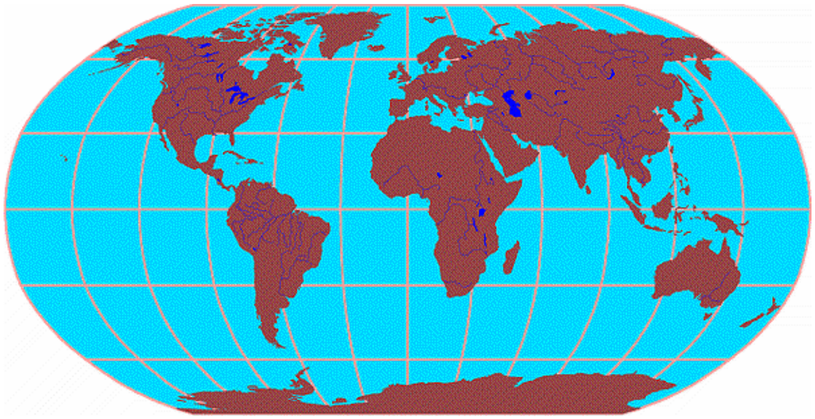
{PRIVATE**}Figure 2a-4:** Gall-Peters projection. Proposed by Arno Peters in 1972, the Gall-Peters projection corrects the distortion of area common in Mercator maps. As a result, it removes the bias in Mercator maps that draws low latitude countries as being smaller than nations in middle and high latitudes. This projection has been officially adopted by a number of United Nations organizations.

The Miller Cylindrical projection is another common two-dimensional map used to represent the entire Earth in a rectangular area (**Figure 2a-5**). In this project, the Earth is mathematically projected onto a cylinder tangent at the equator. This projection in then unrolled to produce a flat two-dimensional representation of the Earth's surface. This projection reduces some of the scale exaggeration present in the Mercator map. However, the Miller Cylindrical projection describes shapes and areas with considerable distortion and directions are true only along the equator.



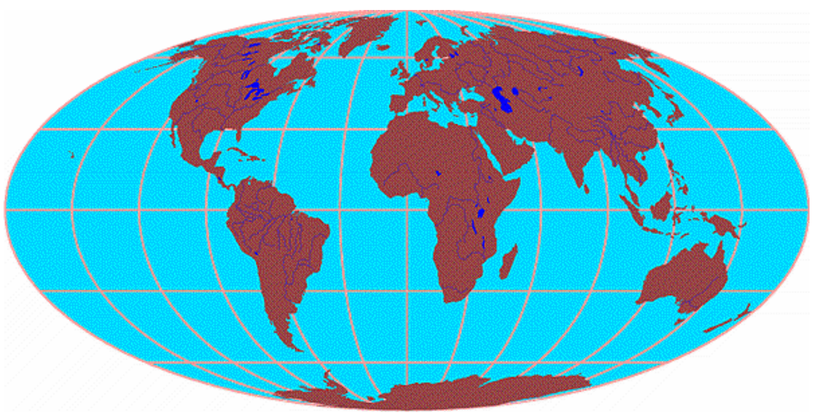
{PRIVATE**}Figure 2a-5:** The Miller Cylindrical projection.

Figure 2a-6 displays the Robinson projection. This projection was developed to show the entire Earth with less distortion of area. However, this feature requires a tradeoff in terms of inaccurate map direction and distance.

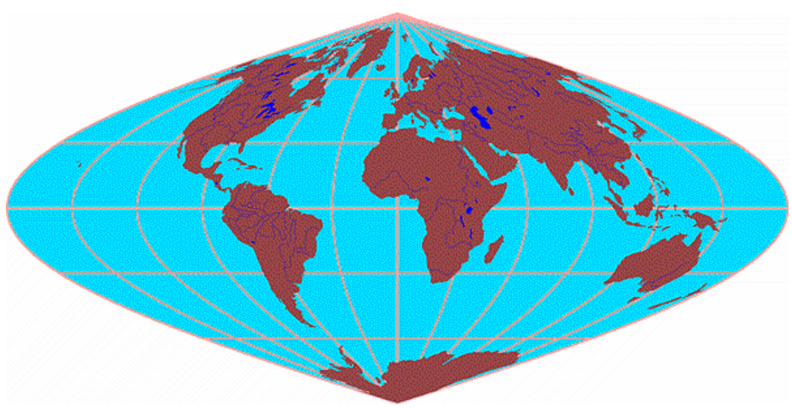


{PRIVATE**}Figure 2a-6:** Robinson's projection. This projection is common in maps that require somewhat accurate representation of area. This map projection was originally developed for Rand McNally and Company in 1961.

The Mollweide projection improves on the Robinson projection and has less area distortion (**Figure 2a-7**). The final projection presented presents areas on a map that are proportional to the same areas on the actual surface of the Earth (**Figure 2a-8**). However, this Sinusoidal Equal-Area projection suffers from distance, shape, and direction distortions.



{PRIVATE**}Figure 2a-7:** Mollweide projection. On this projection the only parallels (line of latitude) drawn of true length are 40° 40' North and South. From the equator to 40° 40' North and South the east-west scale is illustrated too small. From the poles to 40° 40' North and South the east -west scale is illustrated too large.



{PRIVATE**}Figure 2a-8:** Sinusoidal Equal-Area projection.

Map Scale

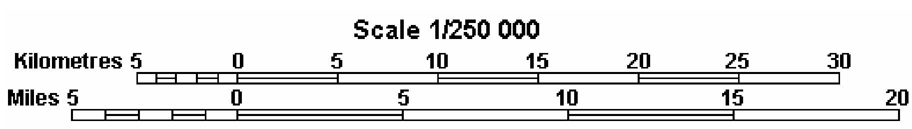
Maps are rarely drawn at the same scale as the real world. Most maps are made at a scale that is much smaller than the area of the actual surface being depicted. The amount of reduction that has taken place is normally identified somewhere on the map. This measurement is commonly referred to as the map scale. Conceptually, we can think of map scale as the ratio between the distance between any two points on the map compared to the actual ground dis tance represented. This concept can also be expressed mathematically as:

$$\text{Map Scale} = \frac{\text{Map Distance}}{\text{Earth Distance}}$$

On most maps, the map scale is represented by a simple fraction or ratio. This type of description of a map's scale is called a representative fraction. For example, a map where one unit (centimeter, meter, inch, kilometer, etc.) on the illustration represents 1,000,000 of these same units on the actual surface of the Earth would have a representative fraction of 1/1,000,000 (fraction) or 1:1,000,000 (ratio). Of these mathematical representations of scale, the ratio form is most commonly found on maps.

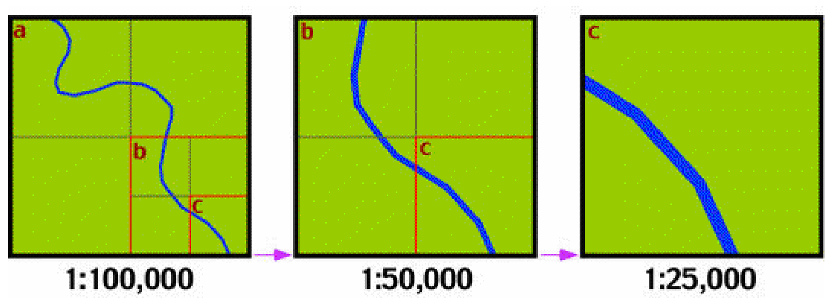
Scale can also be described on a map by a **verbal statement**. For example, 1:1,000,000 could be verbally described as "1 centimeter on the map equals 10 kilometers on the Earth's surface" or "1 inch represents approximately 16 miles".

Most maps also use graphic scale to describe the distance relationships between the map and the real world. In a graphic scale, an illustration is used to depict distances on the map in common units of measurement (**Figure 2a-9**). Graphic scales are quite useful because they can be used to measure distances on a map quickly.



{PRIVATE}**Figure 2a-9:** The following graphic scale was drawn for map with a scale of 1:250,000. In the illustration distances in miles and kilometers are graphically shown.

Maps are often described, in a relative sense, as being either small scale or large scale. **Figure 2a-10** helps to explain this concept. In **Figure 2a-10**, we have maps representing an area of the world at scales of 1:100,000, 1:50,000, and 1:25,000. Of this group, the map drawn at 1:100,000 has the smallest scale relative to the other two maps. The map with the largest scale is map C which is drawn at a scale of 1:25,000.

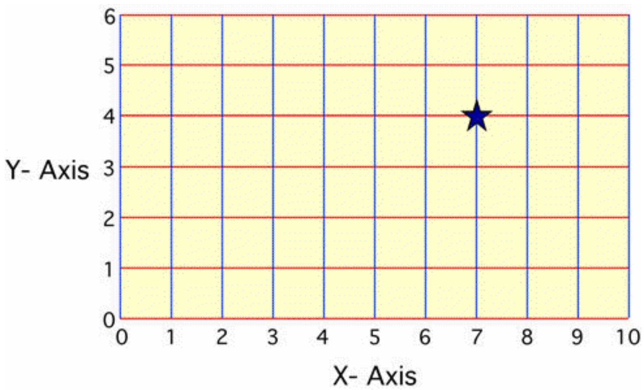


{PRIVATE}**Figure 2a-10:** The following three illustrations describe the relationship between map scale and the size of the ground area shown at three different map scales. The map on the far left has the smallest scale, while the map on the far right has the largest scale. Note what happens to the amount of area represented on the maps when the scale is changed. A doubling of the scale (1:100,000 to 1:50,000 and 1:50,000 to 1:25,000) causes the area shown on the map to be reduced by 25 % or one-quarter.

(b) Location, Distance, and Direction on Maps

{PRIVATE}Location on Maps

Most maps allow us to specify the location of points on the Earth's surface using a coordinate system. For a two-dimensional map, this coordinate system can use simple geometric relationships between the perpendicular axes on a grid system to define spatial location. **Figure 2b-1** illustrates how the location of a point can be defined on a coordinate system.

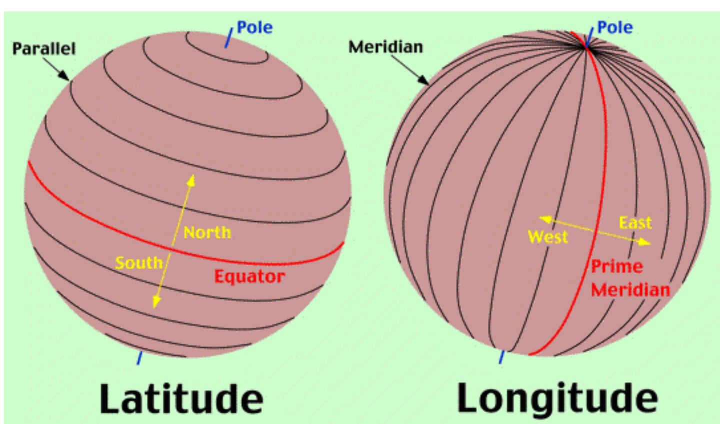


{PRIVATE}Figure 2b-1: A grid coordinate system defines the location of points from the distance traveled along two perpendicular axes from some stated origin. In the example above, the two axes are labeled X and Y. The origin is located in the lower left hand corner. Unit distance traveled along each axis from the origin is shown. In this coordinate system, the value associated with the X-axis is given first, following by the value assigned from the Y-axis. The location represented by the star has the coordinates 7 (X-axis), 4 (Y-axis).

Two types of coordinate systems are currently in general use in geography: the geographical coordinate system and the rectangular (also called **Cartesian**) coordinate system.

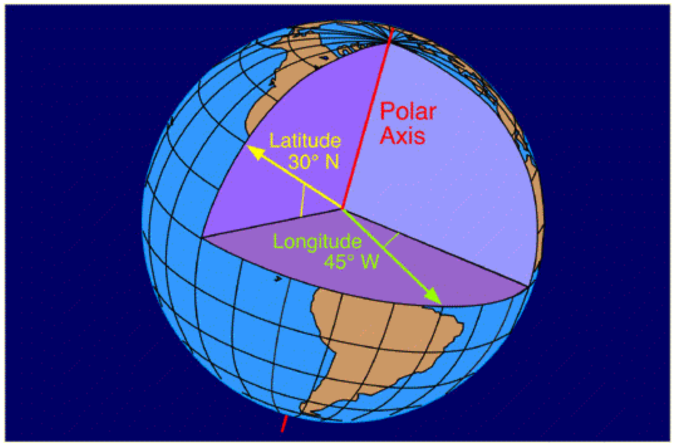
Geographical Coordinate System

The geographical coordinate system measures location from only two values, despite the fact that the locations are described for a three-dimensional surface. The two values used to define location are both measured relative to the polar axis of the Earth. The two measures used in the geographic coordinate system are called latitude and longitude.



{PRIVATE}Figure 2b-2: Lines of latitude or parallels are drawn parallel to the equator (shown in **red**) as circles that span the Earth's surface. These parallels are measure in degrees. There are 90 angular degrees of latitude from the equator to each of the poles. The equator has an assigned value of 0°. Measurements of latitude are also defined as being either north or south of equator to distinguish the hemisphere of their location. Lines of longitude or meridians are circular arcs that meet at the poles. There are 180 degrees of longitude either side of a starting meridian which is known the Prime Meridian. The Prime Meridian has a designated value of 0°. Measurements of longitude are also defined as being either west or east of the Prime Meridian.

Latitude measures the north-south position of locations on the Earth's surface relative to a point found at the center of the Earth (**Figure 2b-2**). This central point is also located on the Earth's rotational or polar axis. The equator is the starting point for the measurement of latitude. The equator has a value of zero degrees. A line of latitude or parallel of 30 degrees North has an angle that is 30 degrees north of the plane represented by the equator (**Figure 2b-3**). The maximum value that latitude can attain is either 90 degrees North or South. These lines of latitude run parallel to the rotational axis of the Earth.

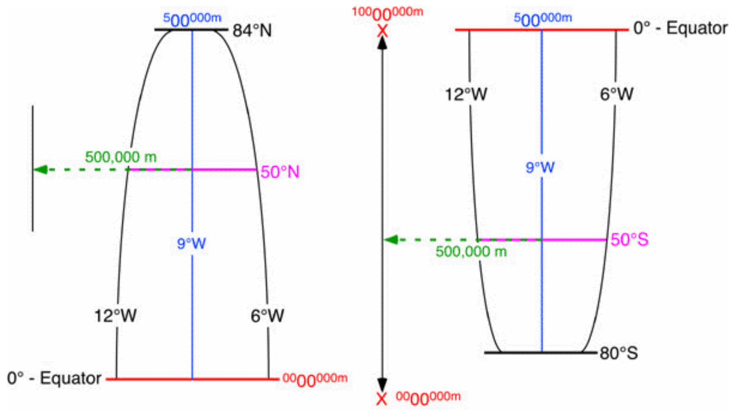


{PRIVATE}**Figure 2b-3:** Measurement of latitude and longitude relative to the equator and the Prime Meridian and the Earth's rotational or polar axis.

Longitude measures the west-east position of locations on the Earth's surface relative to a circular arc called the Prime Meridian (**Figure 2b-2**). The position of the Prime Meridian was determined by international agreement to be in-line with the location of the former astronomical observatory at Greenwich, England. Because the Earth's circumference is similar to circle, it was decided to measure longitude in degrees. The number of degrees found in a circle is 360. The Prime Meridian has a value of zero degrees. A line of longitude or meridian of 45 degrees West has an angle that is 45 degrees west of the plane represented by the Prime Meridian (**Figure 2b-3**). The maximum value that a meridian of longitude can have is 180 degrees which is the distance halfway around a circle. This meridian is called the International Date Line. Designations of west and east are used to distinguish where a location is found relative to the Prime Meridian. For example, all of the locations in North America have a longitude that is designated west.

Universal Transverse Mercator System (UTM)

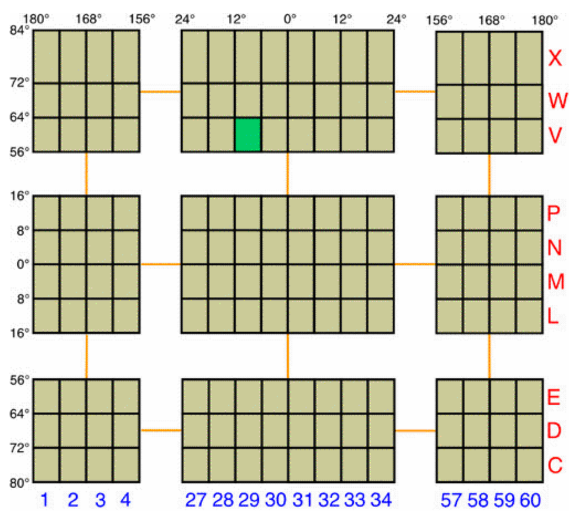
Another commonly used method to describe location on the Earth is the Universal Transverse Mercator (**UTM**) grid system. This rectangular coordinate system is metric, incorporating the meter as its basic unit of measurement. **UTM** also uses the Transverse Mercator projection system to model the Earth's spherical surface onto a two-dimensional plane. The **UTM** system divides the world's surface into 60 - six degree longitude wide zones that run north-south (**Figure 2b-5**). These zones start at the International Date Line and are successively numbered in an eastward direction (**Figure 2b-5**). Each zone stretches from 84 degrees North to 80 degrees South (**Figure 2b-4**). In the center of each of these zones is a central meridian. Location is measured in these zones from a false origin which is determined relative to the intersection of the equator and the central meridian for each zone. For locations in the Northern Hemisphere, the false origin is 500,000 meters west of the central meridian on the equator. Coordinate measurements of location in the Northern Hemisphere using the **UTM** system are made relative to this point in meters in eastings (longitudinal distance) and northings (latitudinal distance). The point defined by the intersection of 50 degrees North and 9 degrees West would have a **UTM** coordinate of Zone **29**, 500000 meters east (E), 5538630 meters north (N) (see **Figures 2b-4** and **2b-5**). In the Southern Hemisphere, the origin is 10,000,000 meters south and 500,000 meters west of the equator and central meridian, respectively. The location found at 50 degrees South and 9 degrees West would have a **UTM** coordinate of Zone **29**, 500000 meters E, 4461369 meters N (remember that northing in the Southern Hemisphere is measured from 10,000,000 meters south of the equator - see **Figures 2b-4** and **2b-5**).



{PRIVATE**}Figure 2b-4:** The following illustration describes the characteristics of the **UTM** zone "**29**" found between 12 to 6 degrees West longitude. Note that the zone has been split into two halves. The half on the left represents the area found in the Northern Hemisphere. The Southern Hemisphere is located on the right. The blue line represents the central meridian for this zone. Locations measurements for this zone are calculated relative to a false origin. In the Northern Hemisphere, this origin is located 500,000 meters west of the equator. The Southern Hemisphere **UTM** measurements are det ermined relative to a origin located at 10,000,000 meters south and 500,000 meters west of the equator and central meridian, respectively.

The UTM system has been modified to make measurements less confusing. In this modification, the six degree wide zones are divided into smaller pieces or quadrilaterals that are eight degrees of latitude tall. Each of these rows is labeled, starting at 80 degrees South, with the letters C to X consecutively with I and O being omitted (**Figure 2b-5**). The last row X differs from the other rows and extends from 72 to 84 degrees North latitude (twelve degrees tall). Each of the quadrilaterals or grid zones are identified by their number/letter designation. In total, 1200 quadrilaterals are defined in the **UTM** system.

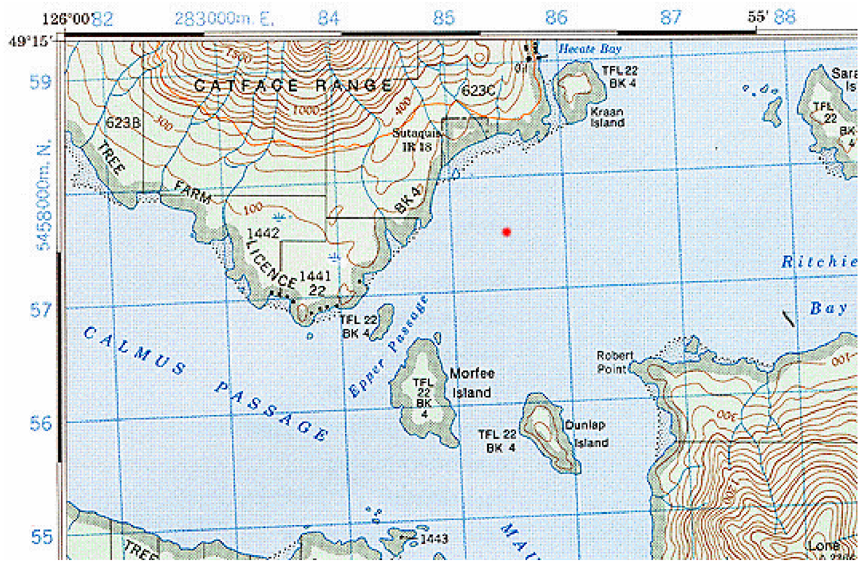
The quadrilateral system allow us to further define location using the **UTM** system. For the location 50 degrees North and 9 degrees West, the **UTM** coordinate can now be expressed as Grid Zone **29U**, 500000 meters E, 5538630 meters N.



{PRIVATE**}Figure 2b-5:** The **UTM** system also uses a grid system to break the Earth up into 1200 quadrilaterals. To keep the illustration manageable, most of these zones have been excluded. Designation of each quadrilaterals is accomplished with a number-letter system. Along the horizontal bottom, the six degree longitude wide zones are numbered, starting at 180 degrees West longitude, from 1 to 60. The twenty vertical rows are assigned letters C to X with I and O excluded. The letter, C, begins at 80 degrees South latitude. Note that the rows are 8 degrees of latitude wide, except for the last row X which is 12 degrees wide. According to the reference system, the bright green quadrilateral has the grid reference **29V** (note that in this system west-east coordinate is given first, followed by the south-north coordinate). This grid zone is found between 56 and 64 degrees North latitude and 6 and 12 degrees West longitude.

Each **UTM** quadrilateral is further subdivided into a number of 100,000 by 100,000 meter zones. These subdivisions are coded by a system of letter combinations where the same two-letter combination is not repeated within 18 degrees of latitude and longitude. Within each of the 100,000 meter squares one can specify location to one-meter accuracy using a 5 digit eastings and northings reference system.

The **UTM** grid system is displayed on all **United States Geological Survey** (**USGS**) and **National Topographic Series** (**NTS**) of Canada maps. On USGS 7.5-minute quadrangle maps (1:24,000 scale), 15-minute quadrangle maps (1:50,000, 1:62,500, and standard-edition 1:63,360 scales), and Canadian 1:50,000 maps the **UTM** grid lines are drawn at intervals of 1,000 meters, and are shown either with blue ticks at the edge of the map or by full blue grid lines. On USGS maps at 1:100,000 and 1:250,000 scale and Canadian 1:250,000 scale maps a full **UTM** grid is shown at intervals of 10,000 meters. **Figure 2b-6** describes how the **UTM** grid system can be used to determine location on a 1:50,000 **National Topographic Series of Canada** map.



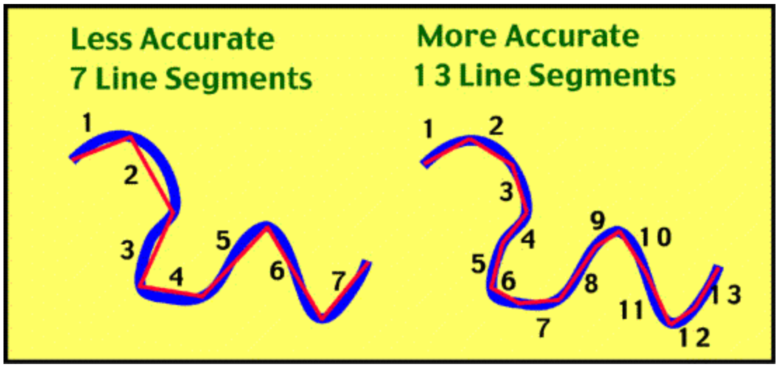
{PRIVATE}**Figure 2b-6:** The top left hand corner the "**Tofino**" 1:50,000 **National Topographic Series of Canada** map is shown above. The blue lines and associated numbers on the map margin are used to determine location by way of the **UTM** grid system. Abbreviated **UTM** 1,000-meter values or **principle digits** are shown by numbers on the map margin that vary from 0 to 100 (100 is actually given the value 00). In each of the corners of the map, two of the principle digits are expressed in their full **UTM** coordinate form. On the image we can see 283000 m E. and 5458000 m N. The **red dot** is found in the center of the grid defined by principle numbers 85 to 86 easting and 57 to 58 northing. A more complete **UTM** grid reference for this location would be 285500 m E. and 5457500 m N. Information found on the map margin also tells us (not shown) that the area displayed is in Grid Zone **10U** and the 100,000 m squares **BK** and **CK** are located on this map.

Distance on Maps

In **section** 2a, w e have learned that depicting the Earth's three-dimensional surface on a two-dimensional map creates a number of distortions that involve distance, area, and direction. It is possible to create maps that are somewhat equidistance. However, even these types of maps have some form of distance distortion. Equidistance maps can only control distortion along either lines of latitude or lines of longitude. Distance is often correct on equidistance maps only in the direction of latitude.

On a map that has a large scale, 1:125,000 or larger, distance distortion is usually insignificant. An example of a large-scale map is a standard topographic map. On these maps measuring straight line distance is simple. Distance is first measured on the map using a ruler. This measurement is then converted into a real world distance using the map's scale. For example, if we measured a distance of 10 centimeters on a map that had a scale of 1:10,000, we would multiply 10 (distance) by 10,000 (scale). Thus, the actual distance in the real world would be 100,000 centimeters.

Measuring distance along map features that are not straight is a little more difficult. One technique that can be employed for this task is to use a number of straight-line segments. The accuracy of this method is dependent on the number of straight-line segments used (**Figure 2b-7**). Another method for measuring curvilinear map distances is to use a mechanical device called an opisometer. This device uses a small rotating wheel that records the distance traveled. The recorded distance is measured by this device either in centimeters or inches.

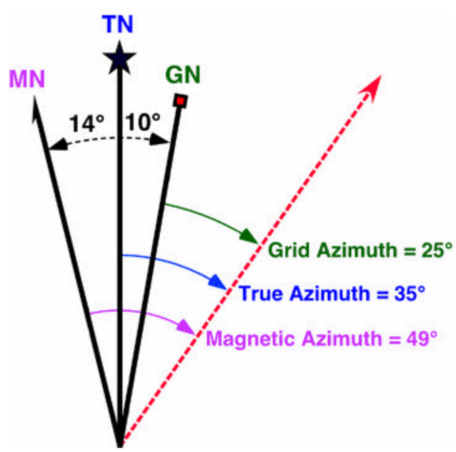


{PRIVATE}**Figure 2b-7:** Measurement of distance on a map feature using straight-line segments.

Direction on Maps

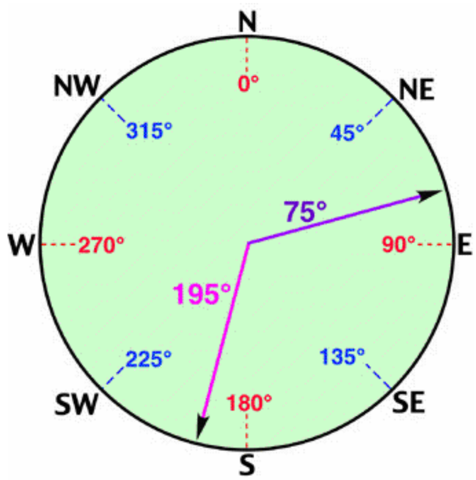
Like distance, direction is difficult to measure on maps because of the distortion produced by projection systems. However, this distortion is quite small on maps with scales larger than 1:125,000. Direction is usually measured relative to the location of North or South Pole. Directions determined from these locations are said to be relative to True North or True South. The magnetic poles can also be used to measure direction. However, these points on the Earth are located in spatially different spots from the geographic North and South Pole. The North Magnetic Pole is located at 78.3 degrees North, 104.0 degrees West near Ellef Ringnes Island, Canada. In the Southern Hemisphere, the South Magnetic Pole is located in Commonwealth Day, Antarctica and has a geographical location of 65 degrees south, 139 degrees east. The magnetic poles are also not fixed overtime and shift their spatial position overtime.

Topographic maps normally have a declination diagram drawn on them (**Figure 2b-8**). On Northern Hemisphere maps, declination diagrams describe the angular difference between Magnetic North and True North. On the map, the angle of True North is parallel to the depicted lines of longitude. Declination diagrams also show the direction of Grid North. Grid North is an angle that is parallel to the easting lines found on the Universal Transverse Mercator (**UTM**) grid system (**Figure 2b-8**).



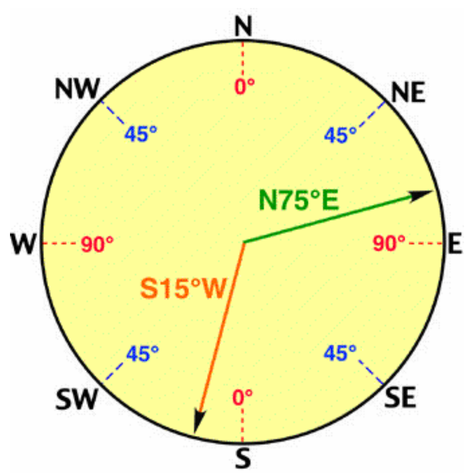
{PRIVATE}**Figure 2b-8:** This declination diagram describes the angular difference between Grid, True, and Magnetic North. This illustration also shows how angles are measured relative grid, true, and magnetic azimuth.

In the field, the direction of features is often determined by a magnetic compass which measures angles relative to Magnetic North. Using the declination diagram found on a map, individuals can convert their field measures of magnetic direction into directions that are relative to either Grid or True North. Compass directions can be described by using either the azimuth system or the bearing system. The azimuth system calculates direction in degrees of a full circle. A full circle has 360 degrees (**Figure 2b-9**). In the azimuth system, north has a direction of either the zero or 360 degrees. East and west have an azimuth of 90 degrees and 270 degrees, respectively. Due south has an azimuth of 180 degrees.



{PRIVATE}**Figure 2b-9:** ***Azimuth*** system for measuring direction is based on the 360 degrees found in a full circle. The illustration shows the angles associated with the major cardinal points of the compass. Note that angles are determined clockwise from north.

The bearing system divides direction into four quadrants of 90 degrees. In this system, north and south are the dominant directions. Measurements are determined in degrees from one of these directions. The measurement of two angles based on this system are described in **Figure 2b-10**.



{PRIVATE}**Figure 2b-10:** The bearing system uses four quadrants of 90 degrees to measure direction. The illustration shows two direction measurements. These measurements are made relative to either north or south. North and south are given the measurement o degrees. East and west have a value of 90 degrees. The first measurement (**green**) is found in the north - east quadrant. As a result, its measurement is north 75 degrees to the east or N75°E. The first measurement (**orange**) is found in the south - west quadrant. Its measurement is south 15 degrees to the west or S15°W.

Global Positioning Systems

Determination of location in field conditions was once a difficult task. In most cases, it required the use of a topographic map and landscape features to estimate location. However, technology has now made this task very simple. Global Positioning Systems (**GPS**) can calculate one's location to an accuracy of about 30-meters (**Figure**

2b-11). These systems consist of two parts: a GPS receiver and a network of many satellites. Radio transmissions from the satellites are broadcasted continually. The GPS receiver picks up these broadcasts and through triangulation calculates the altitude and spatial position of the receiving unit. A minimum of three satellite is required for triangulation.

{PRIVATE**}Figure 2b-11:** Handheld Global Positioning Systems (**GPS**). GPS receivers can determine latitude, longitude, and elevation anywhere on or above the Earth's surface from signals transmitted by a number of satellites. These units can also be used to determine direction, distance traveled, and determine routes of travel in field situations.

(c) Map Location and Time Zones

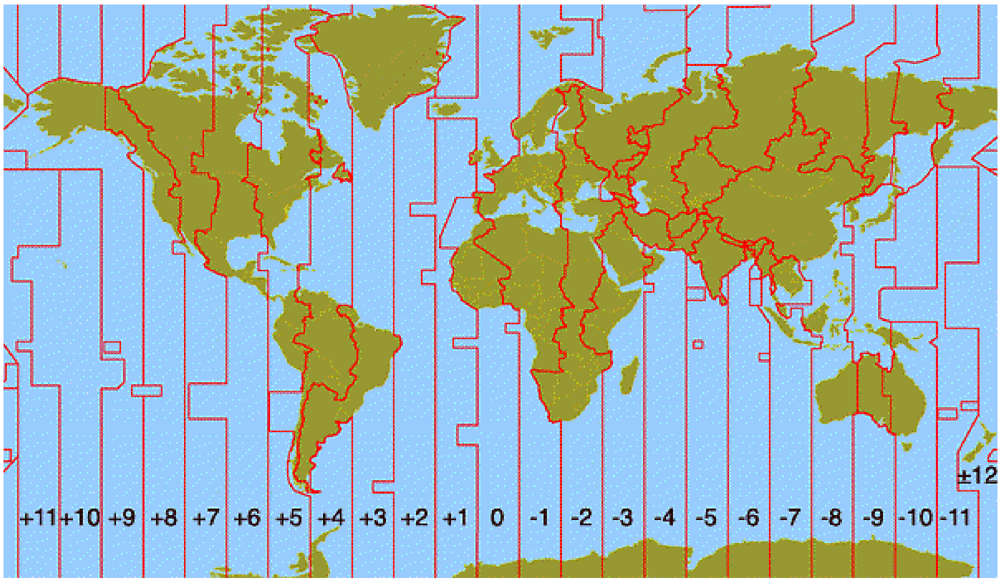
{PRIVATE**}**Before the late nineteenth century, time keeping was essentially a local phenomenon. Each town would set their clocks according to the motions of the sun. Noon was defined as the time when the sun would reached its maximum altitude above the horizon. Cities and towns would assign a clockmaker to calibrate a town clock to these solar motions. This town clock would then represent "official" time and the citizens would set their watches and clocks accordingly.

The later half of the nineteenth century was time of increased movement of humans. In the United States and Canada, large numbers of people were moving west and settlements in these areas began expanding rapidly. To support these new settlements, railroads moved people and resources between the various cities and towns. However, because of the nature of how local time was kept, the railroads experience major problems in constructing timetables for the various stops. Timetables could only become more efficient if the towns and cities adopted some type of standard method of keeping time.

In 1878, Canadian Sir Sanford Fleming suggested a system of worldwide time zones that would simplify the keeping of time across the Earth. Fleming proposed that the globe be divided into 24 time zones, each 15 degrees of longitude in width. Since the world rotates once every 24 hours on its axis and there are 360 degrees of longitude, each hour of Earth rotation represents 15 degrees of longitude.

Railroad companies in Canada and the United States began using Fleming's time zones in 1883. In 1884, an International Prime Meridian Conference was held in Washington D.C. to adopted the standardize method of time keeping and determined the location of the Prime Meridian. Conference members agreed that the longitude of Greenwich, England would become zero degrees longitude and established the 24 time zones relative to the Prime Meridian. It was also proposed that the measurement of time on the Earth would be made relative to the astronomical measurements at the Royal Observatory at Greenwich. This time standard was called Greenwich Mean Time (**GMT**).

Today, many nations operate on variations of the time zones suggested by Sir Fleming. **Figure 2c-1** describes the various time zones currently used on the Earth. In this system, time in the various zones is measured relative the Coordinated Universal Time (**UTC**) standard at the Prime Meridian. **Coordinated Universal Time** became the standard legal reference of time all over the world in 1972. **UTC** is determined from six primary atomic clocks that are coordinated by the **International Bureau of Weights and Measures** (**BIPM**) located in France. The numbers located at the bottom of **Figure 2c-1** indicate how many hours each zone is ahead (negative sign) or behind (positive sign) the **Coordinated Universal Time** standard. Also note that national boundaries and political matters influence the shape of the time zone boundaries. For example, China uses a single time zone (eight hours ahead of **Coordinated Universal Time**) instead of five different time zones.



{PRIVATE**}Figure 2c-1:** Modern standard times zones as measured relative to Coordinated Universal Time. The numbers located at the bottom indicate how many hours each zone is ahead (negative sign) or behind (positive sign) **Coordinated Universal Time**. Some nations (for example, Australia and India) have offset their time zones by half an hour. This situation is not shown on the illustration.

(d) Topographic Maps

{PRIVATE**}Introduction**

A topographic map is a detailed and accurate two-dimensional representation of natural and human-made features on the Earth's surface. These maps are used for a number of applications, from camping, hunting, fishing, and hiking to urban planning, resource management, and surveying. The most distinctive characteristic of a topographic map is that the three-dimensional shape of the Earth's surface is modeled by the use of contour lines. Contours are imaginary lines that connect locations of similar elevation. Contours make it possible to represent the height of mountains and steepness of slopes on a two-dimensional map surface. Topographic maps also use a variety of symbols to describe both natural and human made features such as roads, buildings, quarries, lakes, streams, and vegetation.

Topographic maps produced by the **Canadian National Topographic System** (**NTS**) are generally available in two different scales: 1:50,000 and 1:250,000. Maps with a scale of 1:50,000 are relatively large-scale covering an area approximately 1000 square kilometers. At this scale, features as small as a single home can be shown. The smaller scale 1:250,000 topographic map is more of a general purpose reconnaissance-type map. A map of this scale covers the same area of land as sixteen 1:50,000 scale maps.

In the United States, topographic maps have been made by the United States Geological Survey (**USGS**) since 1879. Topographic coverage of the United States is available at scales of 1:24,000, 1:25,000 (metric), 1:62,250, 1:63,360 (Alaska only), 1:100,000 and 1:250,000.

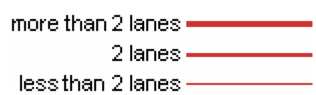
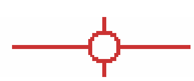
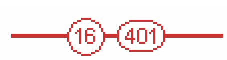
Topographic Map Symbols

Topographic maps use symbols to represent natural and human constructed features found in the environment. The symbols used to represent features can be of three types: points, lines, and polygons. Points are used to depict features like bridges and buildings. Lines are used to graphically illustrate features that are linear. Some common linear features include roads, railways, and rivers. However, we also need to include representations of area, in the case of forested land or cleared land; this is done through the use of color.

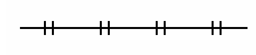
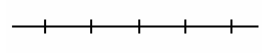
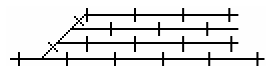
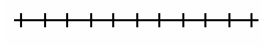
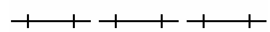
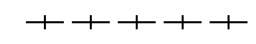
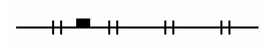
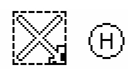
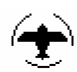
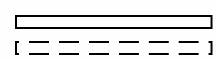
The set of symbols used on **Canadian National Topographic System** (**NTS**) maps has been standardized to simplify the map construction process. A description of the complete set of symbols available can be found in a published guide titled: **Standards and Specifications for Polychrome Maps**. This guide guarantees uniform illustration of surface features on both 1:50 000 and 1:250 000 topographic maps. Despite the existence of this guide, we can find that some topographic maps may use different symbols to depict a feature. This occurs because the symbols used are graphically refined over time - as a result the **Standards and Specifications for Polychrome Maps** guide is always under revision.

The **tables** below describe some of the common symbols used on **Canadian National Topographic System** maps (**source**: Centre for Topographic Information, **Natural Resources Canada**). See the following link for the symbols commonly used on USGS topographic maps.

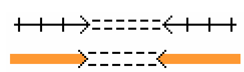
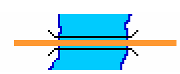
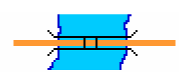
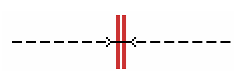
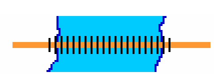
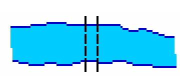
Transportation Features - Roads and Trails

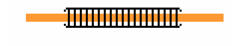
{PRIVATE**}Feature Name**	**Symbol**
Road - hard surface, all season	
Road - hard surface, all season	
Road - loose or stabilized surface, all season	
Road - loose surface, dry weather	
Rapid transit route, road	
Road under construction	
Vehicle track or winter road	
Trail or portage	
Traffic circle	
Highway route number	

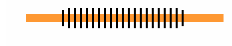
Transportation Features – Railways and Airports

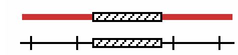
{PRIVATE}**Feature Name**	**Symbol**
Railway - multiple track	
Railway - single track	
Railway sidings	
Railway - rapid transit	
Railway - under construction	
Railway - abandoned	
Railway on road	
Railway station	
Airfield; Heliport	
Airfield, position approximate	
Airfield runways; paved, unpaved	

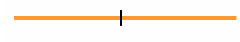
Other Transportation Features - Tunnels, Bridges, etc.

{PRIVATE}**Feature Name**	**Symbol**
Tunnel; railway, road	
Bridge	
Bridge; swing, draw, lift	
Footbridge	
Causeway	
Ford	

Cut 

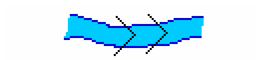
Embankment 

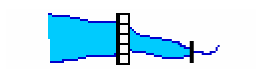
Snow shed 

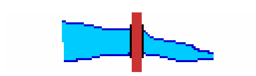
Barrier or gate 

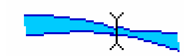
Hydrographic Features - Human Made

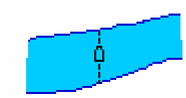
{PRIVATE**}Feature Name** **Symbol**

Lock 

Dam; large, small 

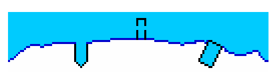
Dam carrying road 

Footbridge 

Ferry Route 

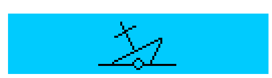
Pier; Wharf; Seawall 

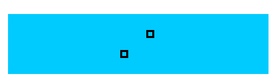
Breakwater 

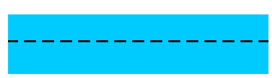
Slip; Boat ramp; Drydock 

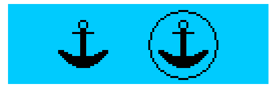
Canal; navigable or irrigation

Canal, abandoned

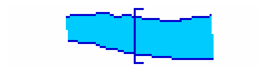
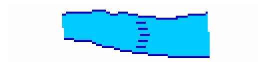
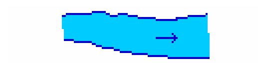
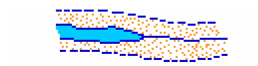
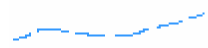
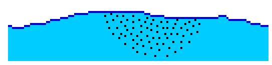
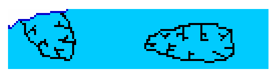
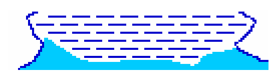
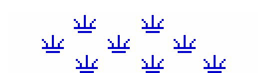
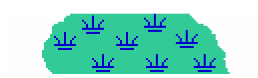
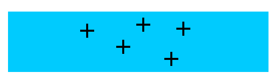
Shipwreck, exposed 

Crib or abandoned bridge pier 

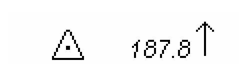
Submarine cable 

Seaplane anchorage; Seaplane base	

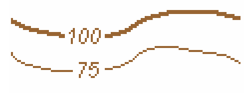
Hydrographic Features - Naturally Occurring

{PRIVATE}**Feature Name**	**Symbol**
Falls	
Rapids	
Direction of flow arrow	
Dry river bed	
Stream - intermittent	
Sand in Water or Foreshore Flats	
Rocky ledge, reef	
Flooded area	
Marsh, muskeg	
Swamp	
Well, water or brine; Spring	
Rocks in water or small islands	
Water elevation	

Terrain Features – Elevation

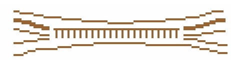
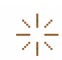
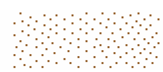
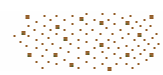
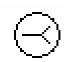
{PRIVATE}**Feature Name**	**Symbol**
Horizontal control point; Bench mark with elevation	
Precise elevation	

Contours; index, intermediate

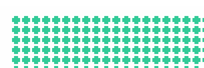


Depression contours

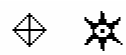
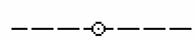
Terrain Features - Geology and Geomorphology

{PRIVATE}Feature Name	**Symbol**
Cliff or escarpment	
Esker	
Pingo	
Sand	
Moraine	
Quarry	
Cave	

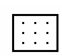
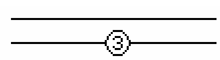
Terrain Features - Land Cover

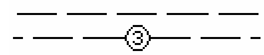
{PRIVATE}Feature Name	**Symbol**
Wooded area	
Orchard	
Vineyard	

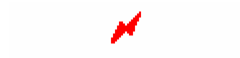
Human Activity Symbols - Recreation

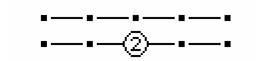
{PRIVATE}**Feature Name**	**Symbol**
Sports track	
Swimming pool	
Stadium	
Golf course	
Golf driving range	
Campground; Picnic site	
Ski area, ski jump	
Rifle range with butts	
Historic site or point of interest; Navigation light	
Aerial cableway, ski lift	

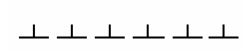
Human Activity Symbols - Agriculture and Industry

{PRIVATE}**Feature Name**	**Symbol**
Silo	
Elevator	
Greenhouse	
Wind-operated device; Mine	
Landmark object (with height); tower, chimney, etc.	
Oil or natural gas facility	
Pipeline, multiple pipelines, control valve	

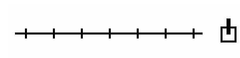
Pipeline, underground multiple pipelines, underground 

Electric facility 

Power transmission line multiple lines 

Telephone line 

Fence

Crane, vertical and horizontal 

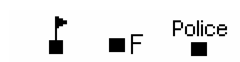
Dyke or levee

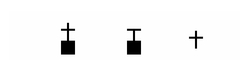
Firebreak

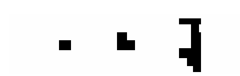
Cut line

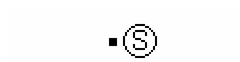
Human Activity Symbols - Buildings

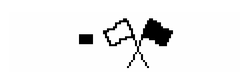
{PRIVATE**}Feature Name** **Symbol**

School; Fire station; Police station 

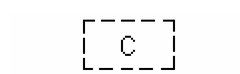
Church; Non-Christian place of worship; Shrine 

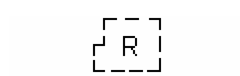
Building 

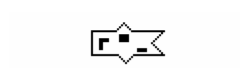
Service centre 

Customs post 

Coast Guard station 

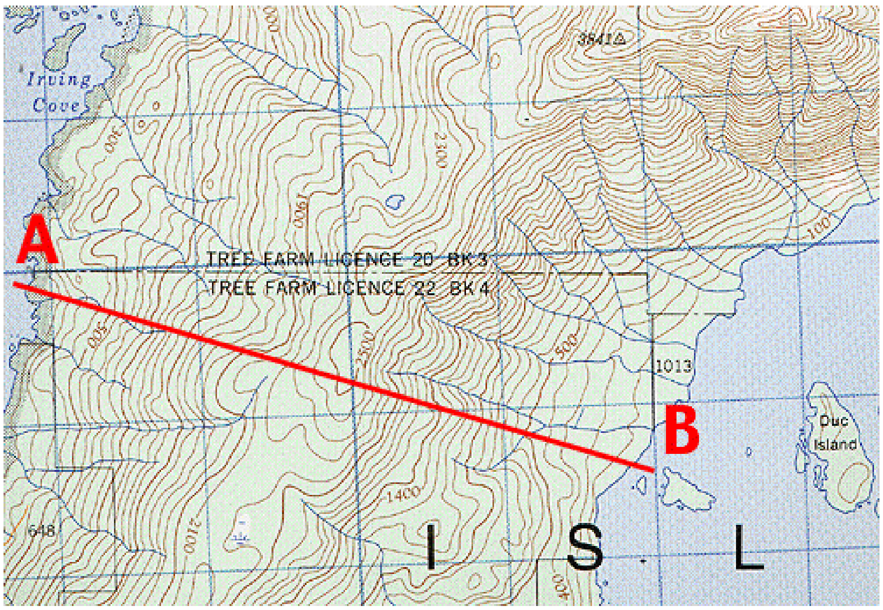
Cemetery 

Ruins 

Fort 

Contour Lines

Topographic maps can describe vertical information through the use of ***contour lines*** (contours). A contour line is an ***isoline*** that connects points on a map that have the same elevation. Contours are often drawn on a map at a uniform vertical distance. This distance is called the ***contour interval***. The map in the **Figure 2d-1** shows contour lines with an interval of 100 feet. Note that every fifth brown contour lines is drawn bold and has the appropriate elevation labeled on it. These contours are called ***index contours***. On **Figure 2d-1** they represent elevations of 500, 1000, 1500, 2000 feet and so on. The interval at which contours are drawn on a map depends on the amount of the relief depicted and the scale of the map.

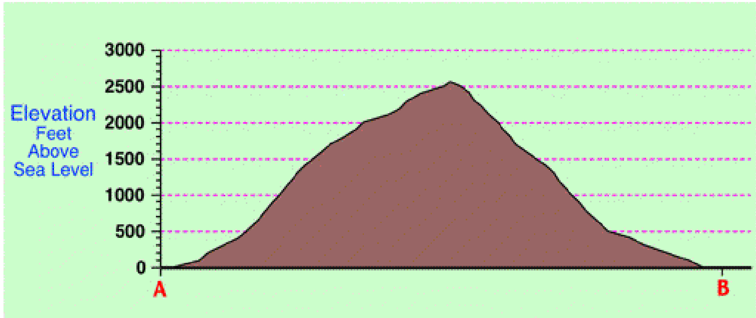


{PRIVATE**}Figure 2d-1:** Portion of the "**Tofino**" 1:50,000 **National Topographic Series of Canada** map. The brown lines drawn on this map are ***contour lines***. Each line represents a vertical increase in elevation of 100 feet. The bold brown contour lines are called ***index contours***. The index contours are labeled with their appropriate elevation which increases at a rate of 500 feet. Note the blue line drawn to separate water from land represents an elevation of 0 feet or ***sea-level***.

Contour lines provide us with a simple effective system for describing landscape configuration on a two-dimensional map. The arrangement, spacing, and shape of the contours provide the user of the map with some idea of what the actual topographic configuration of the land surface looks like. Contour intervals the are spaced closely together describe a steep slope. Gentle slopes are indicated by widely spaced contours. Contour lines that V upwards indicate the presence of a river valley. Ridges are shown by contours that V downwards.

Topographic Profiles

A ***topographic profile*** is a two-dimensional diagram that describes the landscape in vertical cross-section. Topographic profiles are often created from the contour information found on topographic maps. The simplest way to construct a topographic profile is to place a sheet of blank paper along a horizontal transect of interest. From the map, the elevation of the various contours is transferred on to the edge of the paper from one end of the transect to the other. Now on a sheet of graph paper use the x-axis to represent the horizontal distance covered by the transect. The y-axis is used to represent the vertical dimension and measures the change in elevation along the transect. Most people exaggerate the measure of elevation on the y-axis to make changes in relief stand out. Place the beginning of the transect as copied on the piece of paper at the intersect of the x and y-axis on the graph paper. The contour information on the paper's edge is now copied onto the piece of graph paper. **Figure 2d-2** shows a topographic profile drawn from the information found on the transect **A-B** above.

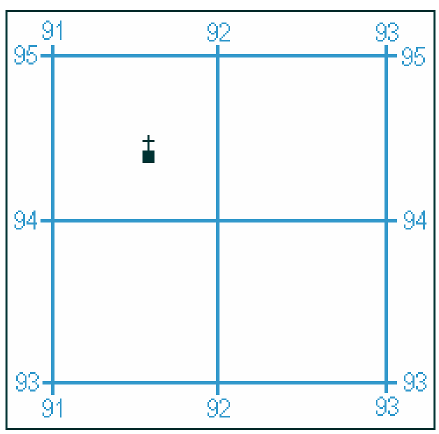


{PRIVATE**}Figure 1d-2:** The following topographic profile shows the vertical change in surface elevation along the transect **AB** from **Figure 1d-1**. A vertical exaggeration of about 4.2 times was used in the profile (horizontal scale = 1:50,000, vertical scale = 1:12,000 and vertical exaggeration = horizontal scale/vertical scale).

Military Grid Reference System and Map Location

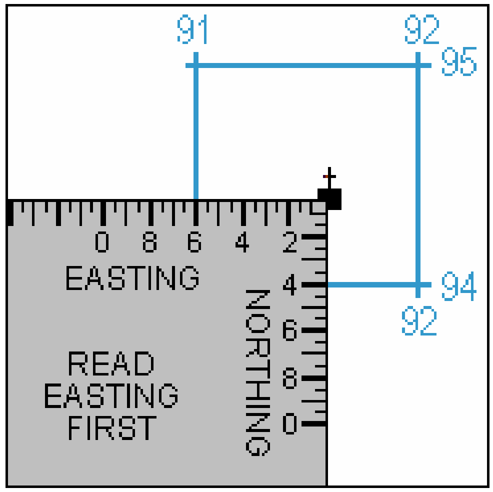
Two rectangular grid systems are available on topographic maps for identifying the location of points. These systems are the ***Universal Transverse Mercator*** **(UTM)** ***grid system*** and the ***Military Grid Reference System***. The **Military Grid Reference System** is a simplified form of **Universal Transverse Mercator grid system** and it provides a very quick and easy method of referencing a location on a topographic map. On a topographic maps with a scale 1:50,000 and larger, the **Military Grid Reference System** is superimposed on the surface of map as blue colored series of equally spaced horizontal and vertical lines. Identifying numbers for each of these lines is found along the map's margin. Each identifying number consists of two digits which range from a value of 00 to 99 (**Figure 2d-3**). Each individual square in the grid system represents a distance of a 1000 by 1000 meters and the total size of the grid is 100,000 by 100,000 meters.

One problem associated with the **Military Grid Reference System** is the fact that reference numbers must be repeat every 100,000 meters. To overcome this difficulty, a method was devised to identify each 100,000 by 100,000 meter grid with two identifying letters which are printed in blue on the border of all topographic maps (note some maps may show more than one grid). When making reference to a location with the **Military Grid Reference System** identifying letters are always given before the horizontal and vertical coordinate numbers.



{PRIVATE**}Figure 2d-3:** Portion of a **Military Grid Reference System** found on a topographic map. Coordinates on this system are based on a **X** (horizontal increasing from left to right) and **Y** (vertical increasing from bottom to top) system. The symbol depicting a church is located in the square **9194**. Note that the value along the **X**-axis (***easting***) is given first followed by the value on the **Y**-axis (***northing***). (**Source**: *Centre for Topographic Information*, **Natural Resources Canada**).

Each individual square in the **Military Grid Reference System** can be further divided into 100 smaller squares (ten by ten). This division allows us to calculate the location of an object to within 100 meters. **Figure 1d-4** indicates that the church is six tenths of the way between lines **91** and **92**, and four tenths of the way between lines **94** and **95**. Using these values, we can state that the ***easting*** as being **916** and the ***northing*** as **944**. By convention, these two numbers are combined into a coordinate reference of **916944**.



{PRIVATE}**Figure 2d-4:** Further determination of the location of the church described in **Figure 2d-3**. Using the calibrated ruler we can now suggest the location of the church to be **916** on the **X**-axis and **944** on the **Y**-axis. Note that the location reference always has an even number of digits, with the three digits representing the ***easting*** and the second three the ***northing***. (**Source**: *Centre for Topographic Information*, **Natural Resources Canada**).

(e) Introduction to Remote Sensing

{PRIVATE}Introduction

Remote sensing can be defined as the collection of data about an object from a distance. Humans and many other types of animals accomplish this task with aid of eyes or by the sense of smell or hearing. Geographers use the technique of remote sensing to monitor or measure phenomena found in the Earth's ***lithosphere***, ***biosphere***, ***hydrosphere***, and ***atmosphere***. Remote sensing of the environment by geographers is usually done with the help of mechanical devices known as ***remote sensors***. These gadgets have a greatly improved ability to receive and record information about an object without any physical contact. Often, these sensors are positioned away from the object of interest by using helicopters, planes, and satellites. Most sensing devices record information about an object by measuring an object's transmission of ***electromagnetic energy*** from ***reflecting*** and ***radiating*** surfaces.

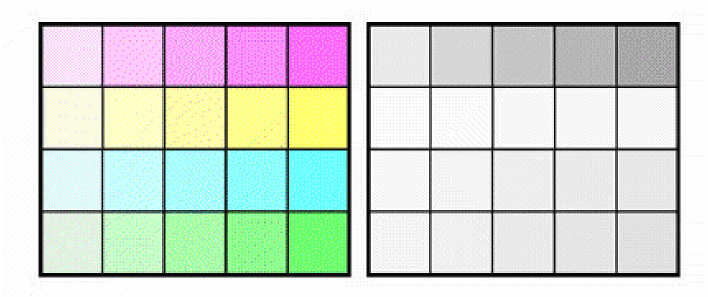
Remote sensing imagery has many applications in mapping land-use and cover, agriculture, soils mapping, forestry, city planning, archaeological investigations, military observation, and geomorphological surveying, among other uses. For example, foresters use aerial photographs for preparing forest cover maps, locating possible access roads, and measuring quantities of trees harvested. Specialized photography using color infrared film has also been used to detect disease and insect damage in forest trees.

The simplest form of remote sensing uses photographic cameras to record information from **visible** or **near *infrared*** wavelengths (**Table 2e-1**). In the late 1800s, cameras were positioned above the Earth's surface in balloons or kites to take ***oblique aerial photographs*** of the landscape. During World War I, ***aerial photography*** played an important role in gathering information about the position and movements of enemy troops. These photographs were often taken from airplanes. After the war, civilian use of aerial photography from airplanes began with the systematic ***vertical*** imaging of large areas of Canada, the United States, and Europe. Many of these images were used to construct topographic and other types of ***reference maps*** of the natural and human-made features found on the Earth's surface.

Table 2e-1: Major regions of the electromagnetic spectrum.

{PRIVATE}**Region Name**	**Wavelength**	**Comments**
Gamma Ray	< 0.03 nanometers	Entirely absorbed by the Earth's atmosphere and not available for remote sensing.
X-ray	0.03 to 30 nanometers	Entirely absorbed by the Earth's atmosphere and not available for remote sensing.
Ultraviolet	0.03 to 0.4 micrometers	Wavelengths from 0.03 to 0.3 micrometers absorbed by ***ozone*** in the Earth's atmosphere.
Photographic Ultraviolet	0.3 to 0.4 micrometers	Available for remote sensing the Earth. Can be imaged with photographic film.
Visible	0.4 to 0.7 micrometers	Available for remote sensing the Earth. Can be imaged with photographic film.
Infrared	0.7 to 100 micrometers	Available for remote sensing the Earth. Can be imaged with photographic film.
Reflected Infrared	0.7 to 3.0 micrometers	Available for remote sensing the Earth. Near Infrared 0.7 to 0.9 micrometers. Can be imaged with photographic film.
Thermal Infrared	3.0 to 14 micrometers	Available for remote sensing the Earth. This wavelength cannot be captured with photographic film. Instead, mechanical sensors are used to image this wavelength band.
Microwave or Radar	0.1 to 100 centimeters	Longer wavelengths of this band can pass through clouds, fog, and rain. Images using this band can be made with sensors that ***actively*** emit microwaves.
Radio	> 100 centimeters	Not normally used for remote sensing the Earth.

The development of color photography following World War II gave a more natural depiction of surface objects. Color aerial photography also greatly increased the amount of information gathered from an object. The human eye can differentiate many more shades of color than tones of gray (**Figure 2e-1** and **2e-2**). In 1942, Kodak developed color infrared film, which recorded wavelengths in the near-infrared part of the electromagnetic spectrum. This film type had good haze penetration and the ability to determine the type and health of vegetation.



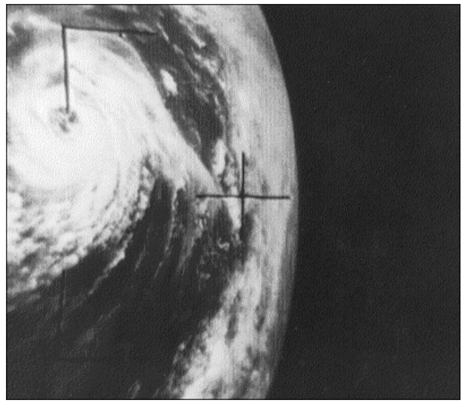
{PRIVATE}**Figure 2e-1:** The rows of color tiles are replicated in the right as complementary gray tones. On the left, we can make out 18 to 20 different shades of color. On the right, only 7 shades of gray can be distinguished.

{PRIVATE}**Figure 2e -2:** Comparison of black and white and color images of the same scene. Note how the increased number of tones found on the color image make the scene much easier to interpret. (**Source:** University of California at Berkley - *Earth Sciences and Map Library*).

Satellite Remote Sensing

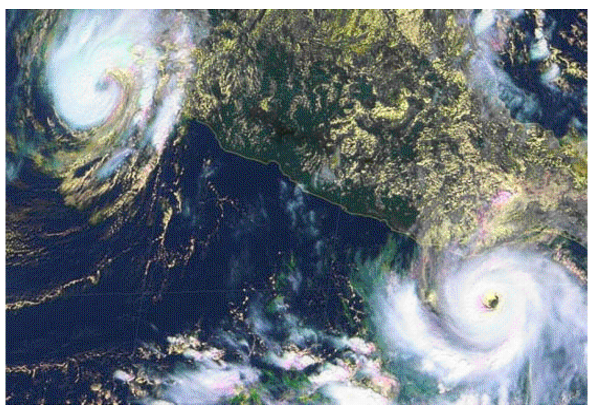
In the 1960s, a revolution in remote sensing technology began with the deployment of space satellites. From their high vantage-point, satellites have a greatly extended view of the Earth's surface. The first meteorological satellite, ***TIROS***-1 (**Figure 2e-3**), was launched by the United States using an Atlas rocket on April 1, 1960. This early weather satellite used vidicon cameras to scan wide areas of the Earth's surface. Early satellite remote sensors did not use conventional film to produce their images. Instead, the sensors digitally capture the images using a device similar to a television camera. Once captured, this data is then transmitted electronically to receiving stations found on the Earth's surface. The image below (**Figure 2e-4**) is from TIROS-7 of a ***mid-latitude cyclone*** off the coast of New Zealand.

{PRIVATE}**Figure 2e -3:** TIROS-1 satellite. (**Source:** NASA - *Remote Sensing Tutorial*).



{PRIVATE**}Figure 2e-4:** TIROS-7 image of a mid-latitude cyclone off the coast of New Zealand, August 24, 1964. (**Source:** NASA - *Looking at Earth From Space*).

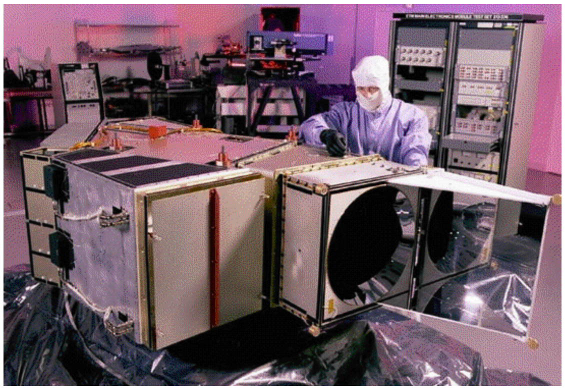
Today, the ***GOES*** (Geostationary Operational Environmental Satellite) system of satellites provides most of the remotely sensed weather information for North America. To cover the complete continent and adjacent oceans two satellites are employed in a ***geostationary orbit***. The western half of North America and the eastern Pacific Ocean is monitored by GEOS-10, which is directly above the equator and 135 degrees West longitude. The eastern half of North America and the western Atlantic are cover by GEOS-8. The GEOS-8 satellite is located overhead of the equator and 75 degrees West longitude. Advanced sensors aboard the GOES satellite produce a continuous data stream so images can be viewed at any instance. The imaging sensor produces visible and infrared images of the Earth's terrestrial surface and oceans (**Figure 2e-5**). Infrared images can depict weather conditions even during the night. Another sounding sensor aboard the satellite provides data to determine vertical temperature profiles, vertical moisture profiles, total ***precipitable water***, and ***atmospheric stability***.



{PRIVATE**}Figure 2e-5:** Color image from GOES-8 of hurricanes Madeline and Lester off the coast of Mexico, October 17, 1998. (**Source:** NASA - *Looking at Earth From Space*).

In the 1970s, the second revolution in remote sensing technology began with the deployment of the ***Landsat*** satellites. Since this 1972, several generations of Landsat satellites with their ***Multispectral Scanners*** (**MSS**) have been providing continuous coverage of the Earth for almost 30 years. Current, Landsat satellites orbit the Earth's surface at an altitude of approximately 700 kilometers. Spatial resolution of objects on the ground surface is 79 x 56 meters. Complete coverage of the globe requires 233 orbits and occurs every 16 days. The Multispectral Scanner records a zone of the Earth's surface that is 185 kilometers wide in four wavelength bands: band 4 at 0.5 to 0.6 micrometers, band 5 at 0.6 to 0.7 micrometers, band 6 at 0.7 to 0.8 micrometers, and band 7 at 0.8 to 1.1 micrometers. Bands 4 and 5 receive the green and red wavelengths in the visible light range of the electromagnetic spectrum. The last two bands image near-infrared wavelengths. A second sensing system was added to Landsat satellites launched after 1982. This imaging system, known as the ***Thematic Mapper***, records seven wavelength bands from the visible to far-infrared portions of the electromagnetic spectrum (**Figure 2e-6**).

In addition, the ground resolution of this sensor was enhanced to 30 x 20 meters. This modification allows for greatly improved clarity of imaged objects.

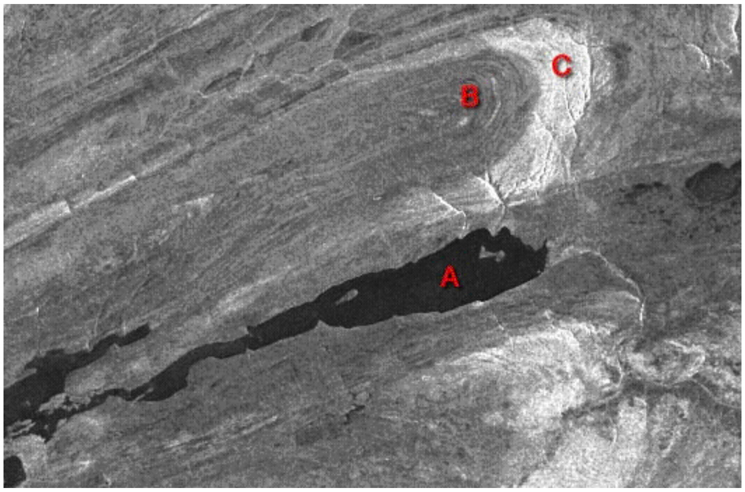


{PRIVATE}**Figure 2e-6:** The Landsat 7 enhanced Thematic Mapper instrument. (**Source:** *Landsat 7 Home Pagehttp://www.spotimage.fr/*).

The usefulness of satellites for remote sensing has resulted in several other organizations launching their own devices. In France, the ***SPOT*** (**Centre National d'Etudes Spatiales**) satellite program has launched four satellites since 1986. Since 1986, SPOT satellites have produced more than 5.5 million images. SPOT satellites use two different sensing systems to image the planet. One sensing system produces black and white panchromatic images from the visible band (0.51-0.73 micrometers) with a ground resolution of 10 x 10 meters. The other sensing device is multispectral capturing green, red, and reflected infrared bands at 20 x 20 meters (**Figure 2d-7**). SPOT-5 will be launched in late 2001. This satellite is much improved from its siblings sporting a ground resolution as low as 2.5 x 2.5 meters in panchromatic mode and 10 x 10 meters in multispectral operation.

{PRIVATE}**Figure 2e-7:** SPOT false-color image of the southern portion of Manhatten Island and part of Long Island, New York. The bridges on the image are (left to right): Brooklyn Bridge, Manhattan Bridge, and the Williamsburg Bridge. (**Source:** *SPOT Image*).

Radarsat-1 was launched by the ***Canadian Space Agency*** in November, 1995. As a remote sensing device, Radarsat is quite different from the Landsat and SPOT satellites. Radarsat is an ***active remote sensing*** system that transmits and receives ***microwave radiation***. Landsat and SPOT sensors ***passively*** measure reflected radiation at wavelengths roughly equivalent to those detected by our eyes. Radarsat's microwave energy penetrates clouds, rain, dust, or haze and produces images regardless of the sun's illumination allowing it to image in darkness. Radarsat images have a resolution between 8 to 100 meters. This sensor has found important applications in crop monitoring, defence surveillance, disaster monitoring, geologic resource mapping, sea-ice mapping and monitoring, oil slick detection, and digital elevation modeling (**Figure 2e-8**).



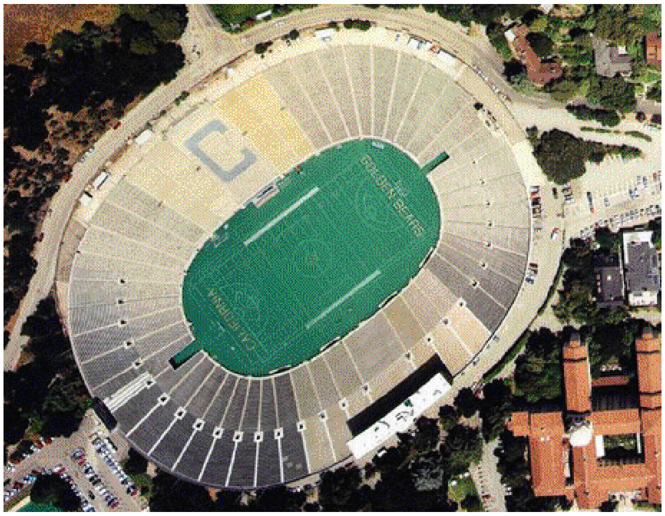
{PRIVATE**}Figure 2e -8:** Radarsat image acquired on March 21, 1996, over Bathurst Island in Nunavut, Canada. This image shows Radarsat's ability to distinguish different types of bedrock. The light shades on this image (**C**) represent areas of limestone, while the darker regions (**B**) are composed of sedimentary siltstone. The very dark area marked **A** is Bracebridge Inlet which joins the Arctic ocean. (**Source:** *Canadian Centre for Remote Sensing - Geological Mapping Bathurst Island, Nunavut, Canada March 21, 1996*).

Principles of Object Identification

Most people have no problem identifying objects from photographs taken from an oblique angle. Such views are natural to the human eye and are part of our everyday experience. However, most remotely sensed images are taken from an overhead or vertical perspective and from distances quite removed from ground level. Both of these circumstances make the interpretation of natural and human-made objects somewhat difficult. In addition, images obtained from devices that receive and capture electromagnetic wavelengths outside human vision can present views that are quite unfamiliar.

To overcome the potential difficulties involved in image recognition, professional image interpreters use a number of characteristics to help them identify remotely sensed objects. Some of these characteristics include:

Shape: this characteristic alone may serve to identify many objects. Examples include the long linear lines of highways, the intersecting runways of an airfield, the perfectly rectangular shape of buildings, or the recognizable shape of an outdoor baseball diamond (**Figure 2e-9**).

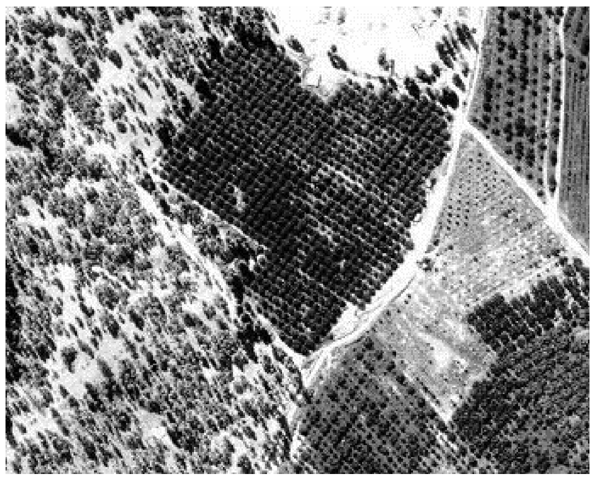


{PRIVATE**}Figure 2e -9:** Football stadium at the University of California at Berkley. Football stadiums have an obvious shape that can be easily recognized even from vertical aerial photographs. (**Source:** University of California at Berkley - *Earth Sciences and Map Library*).

Size: noting the relative and absolute sizes of objects is important in their identification. The scale of the image determines the absolute size of an object. As a result, it is very important to recognize the scale of the image to be analyzed.

Image Tone or Color: all objects reflect or emit specific signatures of electromagnetic radiation. In most cases, related types of objects emit or reflect similar wavelengths of radiation. Also, the types of recording device and recording media produce images that are reflective of their sensitivity to particular range of radiation. As a result, the interpreter must be aware of how the object being viewed will appear on the image examined. For example, on color infrared images vegetation has a color that ranges from pink to red rather than the usual tones of green.

Pattern: many objects arrange themselves in typical patterns. This is especially true of human-made phenomena. For example, orchards have a systematic arrangement imposed by a farmer, while natural vegetation usually has a random or chaotic pattern (**Figure 2e-10**).



{PRIVATE**}Figure 2e -10:** Black and white aerial photograph of natural coniferous vegetation (left) and adjacent apple orchards (center and right).

Shadow: shadows can sometimes be used to get a different view of an object. For example, an overhead photograph of a towering smokestack or a radio transmission tower normally presents an identification problem. This difficulty can be over come by photographing these objects at sun angles that cast shadows. These shadows then display the shape of the object on the ground. Shadows can also be a problem to interpreters because they often conceal things found on the Earth's surface.

Texture: imaged objects display some degree of coarseness or smoothness. This characteristic can sometimes be useful in object interpretation. For example, we would normally expect to see textural differences when comparing an area of grass with a field corn. Texture, just like object size, is directly related to the scale of the image.

(f) Introduction to Geographic Information Systems

{PRIVATE**}Introduction and Brief History**

The advent of cheap and powerful computers over the last few decades has allowed for the development of innovative software applications for the storage, analysis, and display of geographic data. Many of these applications belong to a group of software known as ***Geographic Information Systems*** (**GIS**). Many definitions have been proposed for what constitutes a GIS. Each of these definitions conforms to the particular task that is being performed. Instead of repeating each of these definitions, I would like to broadly define GIS according to what it does. Thus, the activities normally carried out on a GIS include:

The measurement of natural and human made phenomena and processes from a spatial perspective. These measurements emphasize three types of properties commonly associated with these types of ***systems***: ***elements***, ***attributes***, and ***relationships***.

The storage of measurements in digital form in a computer database. These measurements are often linked to features on a digital map. The features can be of three types: points, lines, or areas (polygons).

The analysis of collected measurements to produce more data and to discover new relationships by numerically manipulating and modeling different pieces of data.

The depiction of the measured or analyzed data in some type of display - maps, graphs, lists, or summary statistics.

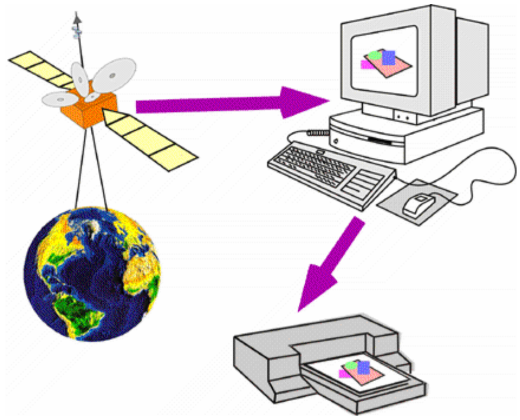
The first computerized GIS began its life in 1964 as a project of the Rehabilitation and Development Agency Program within the government of Canada. The **Canada Geographic Information System** (**CGIS**) was designed to analyze Canada's national land inventory data to aid in the development of land for agriculture. The CGIS project was completed in 1971 and the software is still in use today. The CGIS project also involved a number of key innovations that have found their way into the feature set of many subsequent software developments.

From the mid-1960s to 1970s, developments in GIS were mainly occurring at government agencies and at universities. In 1964, Howard Fisher established the **Harvard Lab for Computer Graphics** where many of the industries early leaders studied. The Harvard Lab produced a number of mainframe GIS applications including: SYMAP (Synagraphic Mapping System),CALFORM, SYMVU, GRID, POLYVRT, and ODYSSEY. ODYSSEY was first modern vector GIS and many of its features would form the basis for future commercial applications. Automatic Mapping System was developed by the United States Central Intelligence Agency (**CIA**) in the late 1960s. This project then spawned the CIA's **World Data Bank**, a collection of coastlines, rivers, and political boundaries, and the **CAM** software package that created maps at different scales from this data. This development was one of the first systematic map databases. In 1969, Jack Dangermond, who studied at the Harvard Lab for Computer Graphics, co-founded ***Environmental Systems Research Institute*** (**ESRI**) with his wife Laura. **ESRI** would become in a few years the dominate force in the GIS marketplace and create ***ArcInfo*** and ***ArcView*** software. The first conference dealing with GIS took place in 1970 and was organized by Roger Tomlinson (key individual in the development of **CGIS**) and Duane Marble (professor at Northwestern University and early GIS innovator). Today, numerous conferences dealing with GIS run every year attracting thousands of attendants.

In the 1980s and 1990s, many GIS applications underwent substantial evolution in terms of features and analysis power. Many of these packages were being refined by private companies who could see the future commercial potential of this software. Some of the popular commercial applications launched during this period include: ***ArcInfo***, ***ArcView***, ***MapInfo***, ***SPANS GIS***, ***PAMAP GIS***, ***INTERGRAPH***, and ***SMALLWORLD***. It was also during this period that many GIS applications moved from expensive minicomputer workstations to personal computer hardware.

Components of a GIS

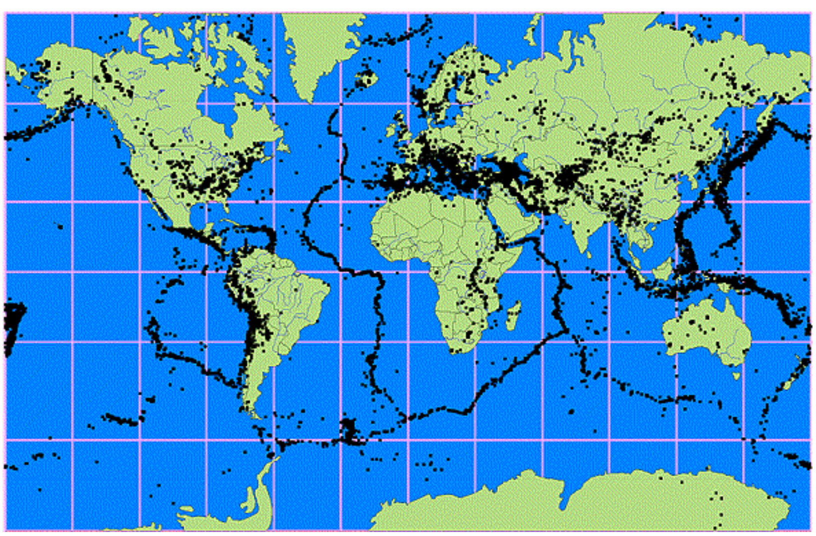
A **Geographic Information System** combines computer cartography with a database management system. **Figure 2f-1** describes some of the major components common to a GIS. This diagram suggests that a GIS consists of three subsystems: (1) an input system that allows for the collection of data to be used and analyzed for some purpose; (2) computer hardware and software systems that store the data, allow for data management and analysis, and can be used to display data manipulations on a computer monitor; (3) an output system that generates hard copy maps, images, and other types of output.



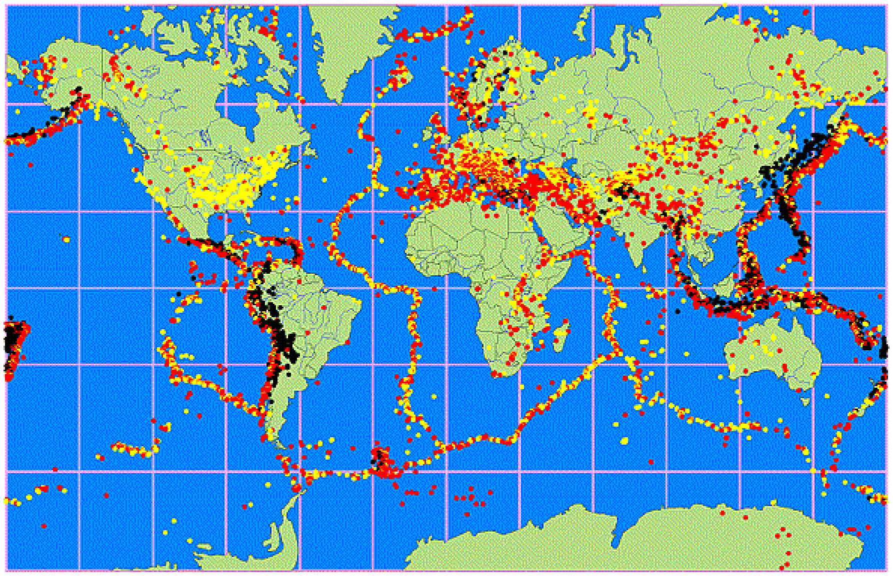
{PRIVATE**}Figure 2f-1:** Three major components of a Geographic Information System. These components consist of input, computer hardware and software, and output subsystems.

Two basic types of data are normally entered into a GIS. The first type of data consists of real world phenomena and features that have some kind of spatial dimension. Usually, these data ***elements*** are depicted mathematically in the GIS as either points, lines, or polygons that are referenced geographically (or ***geocoded***) to some type of coordinate system. This type data is entered into the GIS by devices like scanners, digitizers, GPS, air photos, and satellite imagery. The other type of data is sometimes referred to as an ***attribute***. Attributes are pieces of data that are connected or related to the points, lines, or polygons mapped in the GIS. This attribute data can be analyzed to determine patterns of importance. Attribute data is entered directly into a database where it is associated with element data.

The difference between ***element*** and ***attribute*** data can be illustrated in **Figures 2f-2** and **2f-3**. **Figure 2f-2** shows the location of some of the ***earthquakes*** that have occurred in the last century. These plotted data points can be defined as elements because their main purpose is to describe the location of the earthquakes. For each of the earthquakes plotted on this map, the GIS also has data on their depth. These measurements can be defined as attribute data because they are connected to the plotted earthquake locations in **Figure 2f-2**. **Figure 2f-3** shows the attribute earthquake depth organized into three categories: shallow; intermediate; and deep. This analysis indicates a possible ***relationship*** between earthquake depth and spatial location - deep earthquakes do not occur at the ***mid-oceanic ridges***.

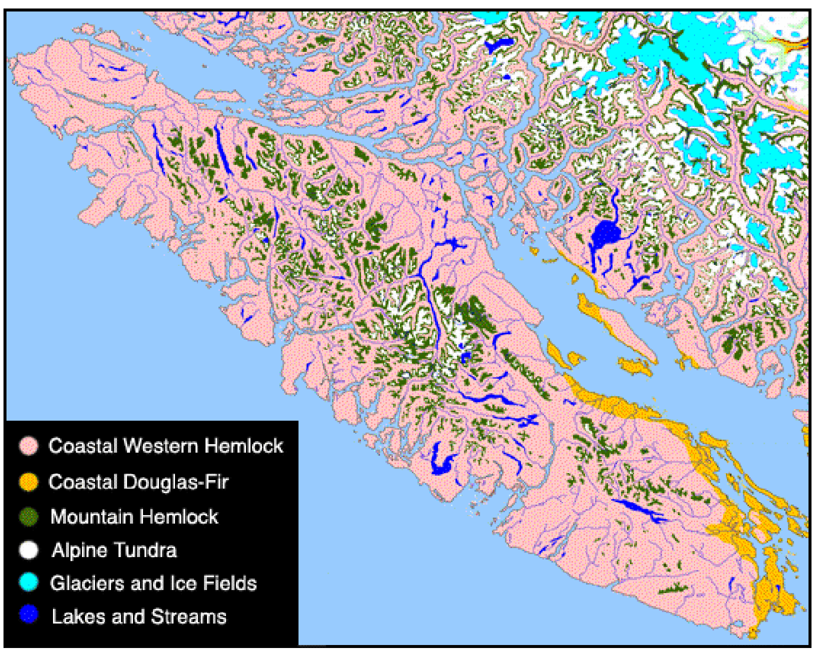


{PRIVATE**}Figure 2f-2:** Distribution of earthquake events that have occurred over the last century.



{PRIVATE**}Figure 2f-3:** Earthquake events organized according to depth (**yellow** (shallow) = surface to 25 kilometers below the surface, **red** (intermediate) = 26 to 75 kilometers below the surface, and **black** (deep) = 76 to 660 kilometers below the surface).

Within the **GIS** database a user can enter, analyze, and manipulate data that is associated with some spatial element in the real world. The cartographic software of the GIS enables one to display the geographic information at any scale or projection and as a variety of layers which can be turned on or off. Each layer would show some different aspect of a place on the Earth. These layers could show things like a road network, topography, vegetation cover, streams and water bodies, or the distribution of annual precipitation received. The output illustrated in **Figure 2f-4** merges data layers for vegetation community type, glaciers and ice fields, and water bodies (streams, lakes, and ocean).



{PRIVATE**}Figure 2f-4:** Graphic output from a GIS. This GIS contains information about the major plant communities, lakes and streams, and glaciers and ice fields found occupying the province of British Columbia, Canada. The output shows Vancouver Island and part of the British Columbia mainland.

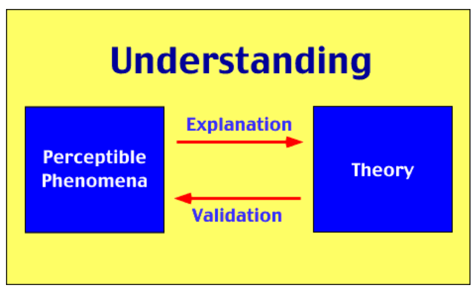
3) The Science of Physical Geography

(a) Scientific Method

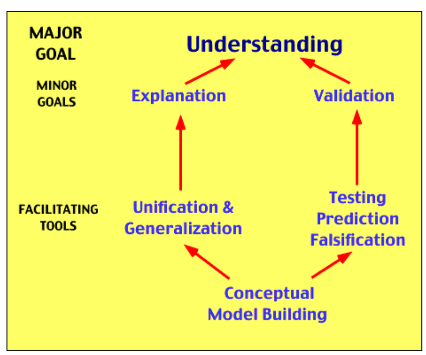
{PRIVATE}***Francis Bacon*** (1561-1626), a 17th century English philosopher, was the first individual to suggest a universal methodology for ***science***. Bacon believed that scientific method required an ***inductive*** process of inquiry. ***Karl Popper*** later refuted this idea in the 20th century. Popper suggested that science could only be done using a ***deductive*** methodology. The next section (***3b***) examines Karl Popper's recommended methodology for doing science more closely.

Science is simply a way of acquiring knowledge about nature and the Universe. To do science, one must follow a specific universal methodology. The central theme of this methodology is the testing of ***hypotheses***. The overall goal of science is to better understand the world around us. Various fields of study, like Physics, Chemistry, Biology, Medicine and the Earth Sciences, have used science exclusively to expand their knowledge base. Science allows its practitioners to acquire knowledge using techniques that are both **neutral** and **unbiased**.

The broadest, most inclusive goal of science is to **understand** (see **Figure 3a-1**). Understanding encompasses a number of other goals of science, many of which are quite specialized (see **Figure 3a-2**). **Explanation** is perhaps the most important basic goal of understanding. Explanation consists of relating observed reality to a system of concepts, laws, or empirically based generalizations. Explanation may also relate observed phenomena to a system of causes, or link them to mechanisms that are hierarchically structured at lower-levels of function.



{PRIVATE}**Figure 3a-1:** Relationship between phenomena, theory, and understanding using scientific method. The interaction between observable phenomena and theory is accomplished through explanation and validation.



{PRIVATE}**Figure 3a-2:** Processes involved in understanding natural phenomena using scientific method.

The secondary goal of **explanation** has two important components: **generalization** and **unification** (see **Figure 3a-2**). **Generalization** may be considered to be the **condensation** of a body of empirical fact into a simple statement. In the process of such condensation, it is likely that some detail must be omitted and the processes and phenomenon abstracted. Generalization may also involve **isolating** the phenomenon from other aspects of the system of interest. This is sometimes referred to as idealization. A second aspect of explanation is the **unification** of apparently unrelated phenomena in the same abstract or ideal system of concepts.

Another minor goal of science is the **validation** of constructed models (**conceptual model building**) of **understanding**. **Validation** is accomplished through ***hypothesis testing***, ***prediction***, and ***falsification***. The next **section** (***3b***) examines these aspects of science in greater detail.

(b) The Hypothetico-Deductive Method

{PRIVATE**}**The philosopher ***Karl Popper*** suggested that it is impossible to prove a scientific theory true by means of ***induction***, because no amount of evidence assures us that contrary evidence will not be found. Instead, Karl Popper proposed that proper science is accomplished by ***deduction***. Deduction involves the process of ***falsification***. Falsification is a particular specialized aspect of ***hypothesis testing***. It involves stating some output from theory in specific and then finding contrary cases using experiments or observations. The methodology proposed by Popper is commonly known as the **hypothetico-deductive method**.

Popper's version of scientific method first begins with the postulation of a ***hypothesis***. A hypothesis is an educated guess or a theory that explains some phenomenon. The researcher then tries to prove or test this scientific theory false through ***prediction*** or ***experimentation*** (see ***Figure 3a-2***).A prediction is a forecast or extrapolation from the current state of the ***system*** of interest. Predictions are most useful if they can go beyond simple forecast. An experiment is a controlled investigation designed to evaluate the outcomes of causal manipulations on some system of interest.

To get a better understanding of the **hypothetico-deductive method**, we can examine the following geographic phenomena. In the brackish tidal marshes of the Pacific Coast of British Columbia and Washington, we find that the plants in these communities spatially arrange themselves in zones that are defined by elevation. Near the shoreline plant communities are dominated primarily by a single species known as Scirpus americanus. At higher elevations on the tidal marsh Scirpus americanus disappears and a species called Carex lyngbyei becomes widespread. The following hypothesis has been postulated to explain this unique phenomenon:

The distribution of Scirpus americanus and Carex lyngbyei is controlled by their tolerances to the frequency of tidal flooding. Scirpus americanus is more tolerant of tidal flooding than Carex lyngbyei and as a result it occupies lower elevations on the tidal marsh. However, Scirpus americanus cannot survive in the zone occupied by Carex lyngbyei because not enough flooding occurs. Likewise, Carex lyngbyei is less tolerant of tidal flooding than Scirpus americanus and as a result it occupies higher elevations on the tidal marsh. Carex lyngbyei cannot survive in the zone occupied by Scirpus americanus because too much flooding occurs.

According to Popper, to test this theory a scientist would now have to prove it false. As discussed above this can be done in two general ways: 1) predictive analysis; or 2) by way of experimental manipulation. Each of these methods has been applied to this problem and the results are described below.

Predictive Analysis

If the theory is correct, we should find that in any tidal marsh plant community that contains Scirpus americanus and Carex lyngbyei that the spatial distribution of these two species should be similar in all cases. This is indeed true. However, there could be some other causal factor, besides flooding frequency, that may be responsible for these unique spatial patterns.

Experimental Manipulation

If the two species are transplanted into the zones of the other they should not be able to survive. An actual transplant experiment found that Scirpus americanus can actually grow in the zone occupied by Carex lyngbyei, while Carex lyngbyei could also grow at lower Scirpus sites. However, this growth became less vigorous as the elevation became lower and at a certain elevation it could not grow at all. These results falsify the postulated theory. So the theory must be modified based on the results and tested again.

The process of testing theories in science is endless. Part of this problem is related to the complexity of nature. Any one phenomenon in nature is influenced by numerous factors each having its particular cause and effect. For this reason, one positive test result is not conclusive proof that the phenomenon under study is explained. However, some tests are better than others and provide us with stronger confirmation. These tests usually allow for the isolation of the phenomena from the effects of causal factors. Manipulative experiments tend to be better than tests based on prediction in this respect.

(c) Concepts of Time and Space in Physical Geography

{PRIVATE**}**The concepts of ***time*** and ***space*** are very important for understanding the function of phenomena in the natural world. Time is important to Physical Geographers because the spatial patterns they study can often only be explained in **historic** terms. The measurement of time is not **absolute**. Time is perceived by humans in a **relative** fashion by using human created units of measurement. Examples of human created units of time are the measurement of seconds, minutes, hours, and days.

Geographers generally conceptualize two types of space. ***Concrete space*** represents the real world or environment. ***Abstract space*** models reality in a way that distills much of the spatial information contained in the real world. Maps are an excellent example of abstract space. Finally, like time, space is also perceived by humans in a **relative** fashion by using human created units of measurement.

Both time and space are variable in terms of ***scale***. As such, researchers of natural phenomena must investigate their subjects in the appropriate temporal and/or spatial scales. For example, an investigator studying a forest ecosystem will have to deal with completely different scales of time and space when compared to a researcher examining soil bacteria. The trees that make up a forest generally occupy large tracts of land. For example, the boreal forest occupies millions of hectares in Northern Canada and Eurasia. Temporally, these trees have life spans that can be as long as several hundred years. On the other hand, soil bacteria occupy much smaller spatial areas and have life spans that can be measured in hours and days.

(d) Study of Form or Process?

{PRIVATE**}**Physical Geography as a science is experiencing a radical change in philosophy. It is changing from a science that was highly **descriptive** to one that is increasingly **experimental** and **theoretical**. This transition represents a strong desire by Physical Geographers to understand the **processes** that cause the patterns or **forms** we see in nature.

Before 1950, the main purpose of research in Physical Geography was the description of the natural phenomena. Much of this description involved measurement for the purpose of gaining basic facts dealing with **form** or spatial appearance. Out of this research Physical Geographers determined such things as: the climatic characteristics for specific locations and regions of the planet; flow rates of rivers; soil characteristics for various locations on the Earth's surface; distribution ranges of plant and animal species; and calculations of the amount of freshwater stored in lakes, glaciers, rivers and the atmosphere. By the beginning of the 20th century Physical Geographers began to examine the descriptive data that was collected, and started to ask questions related to why? Why is the climate of urban environments different from the climate of rural? Why does hail only form in thunderstorms? Why are soils of the world's tropical regions nutrient poor? Why do humid and arid regions of the world experience different levels of erosion?

In Physical Geography, and all other sciences, most questions that deal with why are usually queries about **process**. Some level of understanding about process can be derived from basic descriptive data. Process is best studied, however, through experimental manipulation and hypothesis testing. By 1950, Physical Geographers were more interested in figuring out process than just collecting descriptive facts about the world. This attitude is even more prevalent today because of our growing need to understand how humans are changing the Earth and its environment.

Finally, as mentioned above, a deeper understanding of process normally requires the use of hypothesis testing, experimental methods, and statistics. As a result, the standard undergraduate and graduate curriculum in Physical Geography exposes students to this type of knowledge so they can better ask the question why?

(e) Descriptive Statistics

{PRIVATE**}Introduction**

Physical Geographers often collect quantitative information about natural phenomena to further knowledge in their field of interest. This collected data is then often analyzed statistically to provide the researcher with impartial and enlightening presentation, summary, and interpretation of the phenomena understudy. The most common statistical analysis performed on data involves the determination of descriptive characteristics like measures of central tendency and dispersion.

It usually is difficult to obtain measurements of all the data available in a particular system of interest. For example, it may be important to determine the average atmospheric pressure found in the center of hurricanes. However, to make a definitive conclusion about a hurricane's central pressure with 100 % confidence would require the measuring of all the hurricanes that ever existed on this planet. This type of measurement is called a ***population parameter***. Under normal situations, the determination of population parameters is impossible, and we settle with a subset measure of the population commonly called an ***estimator***. Estimators are determined by taking a representative ***sample*** of the population being studied.

Samples are normally taken at ***random***. Random sampling implies that each measurement in the population has an equal chance of being selected as part of the sample. It also ensures that the occurrence of one measurement in a sample in no way influences the selection of another. Sampling methods are **biased** if the recording of some influences the recording of others or if some members of the population are more likely to be recorded than others.

Measures of Central Tendency

Collecting data to describe some phenomena of nature usually produces large arrays of numbers. Sometimes it is very useful to summarize these large arrays with a single parameter. Researchers often require a summary value that determines the center in a data sample's distribution. In orther words, a measure of the central tendency of the data set. The most common of these measures are the ***mean***, the ***median***, and the ***mode***.

Table 3e-1 describes a 15-year series of number of days with precipitation in December for two fictitious locations. The following discussion describes the calculation of the ***mean***, ***median***, and ***mode*** for this ***sample*** data set.

Table 3e-1: Number of days with precipitation in December for **Piney** and **Steinback**, 1967-81.

{PRIVATE**} Year**	**Piney**	**Steinback**
1967	10	12
1968	12	12
1969	9	13
1970	7	15
1971	10	13
1972	11	9
1973	9	16
1974	10	11
1975	9	12
1976	13	13
1977	8	10
1978	9	9
1979	10	13
1980	8	14
1981		
1981	9	15
Σ (Xi)	144	187
N	15	15

The ***mean*** values of these two sets is determined by summing of the yearly values divided by the number of observations in each data set. In mathematical notation this calculation would be expressed as:

mean (□) = Σ(Xi)/N where Xi is the individual values,

N is the number of values, and □ is **sigma**, a sign used to show summation.

Thus, the calculate means for **Piney** and **Steinback** are:

Piney *mean* = 10 (rounded off)

Steinback *mean* = 13 (rounded off)

The ***mode*** of a data series is that value that occurs with **greatest frequency**. For Piney, the most frequent value is 9 which occurs five times. The mode for Steinback is 13.

The third measure of central tendency is called the ***median***. The median is the middle value (or the average of the two middle values in an even series) of the data set when the observations are organized in ascending order. For the two locations in question, the medians are:

Piney

9, 9, 10, 11, 12, 12, 12, **13**, 13, 13, 13, 14, 15, 15, 16 ***median*** = 13

Steinback

7, 8, 8, 9, 9, 9, 9, **9**, 10, 10, 10, 10, 11, 12, 13 ***median*** = 9

Measures of Dispersion

Measures of central tendency provide no clue into how the observations are dispersed within the data set. Dispersion can be calculated by a variety of descriptive statistics including the ***range***, ***variance***, and ***standard deviation***. The simpest measure of dispersion is the ***range***. The ***range*** is calculated by subtracting the smallest individual value from the largest. When presented together with the ***mean***, this statistic provides a measure of data set variability. The ***range***, however, does not provide any understanding to how the data are distributed about the ***mean***. For this measurement, the ***standard deviation*** is of value.

The following information describes the calculation of the ***range***, ***variance***, and standard deviation for the data set in **Table 3e-2**.

Table 3e-2: Dates of the first fall frost at Somewhere, USA, for an 11-year period.

{PRIVATE}Day of First Frost * (Xi)	Xi - $\bar{X}$	(Xi - $\bar{X}$)^2
291	-8	64
299	0	0
279	-20	400
302	3	9
280	-19	361
303	4	16
299	0	0
304	5	25
307	8	64
314	15	225
313	14	196
Σ(Xi) = 3291 $\bar{X}$ = 3291/11 = 299		Σ(Xi - $\bar{X}$)^2 = 1360

{PRIVATE}*The dates are given in year days, i.e., January 1st is day 1, January 2nd is day 2, and so on throughout the year.

The ***range*** for the data is set is derived by subtracting 279 (the smallest value) from 314 (the largest value). The ***range*** is 35 days.

The first step in the calculation of ***standard deviation*** is to determine the ***variance*** by obtaining the deviations of the individual values (Xi) from the mean ($\overline{X}$). The formula for ***variance*** ($\mathbf{S^2}$) is:

$\mathbf{S^2} = [\Sigma(Xi - \overline{X})^2] / (N-1)$

where Σ is the summation sign, $(Xi - \overline{X})^2$ is calculated (third column), and **N** is the number of observations. ***Standard deviation*** (**S**) is merely the **square root** of the ***variance*** ($\sqrt{\mathbf{S^2}}$) **.**

In the case of the Somewhere data, the ***standard deviation*** is:

$\mathbf{S^2} = 1356 / 10$

S = 11.6 or 12 (to the nearest day)

This value provides significant information about the distribution of data around the mean. For example:

(a) The ***mean*** ± one sample ***standard deviation*** contains approximately 68 % of the measurements in the data series.

(b) The ***mean*** ± two sample ***standard deviations*** contains approximately 95 % of the measurements in the data series.

In Somewhere, the corresponding dates for fall frosts ± one and two ***standard deviations*** from the ***mean*** (day 299) are:

Minus two standard deviations: 299 - 24 = 275

Minus one standard deviation: 299 - 12 = 287

Plus one standard deviation: 299 + 12 = 311

Plus two standard deviations: 299 + 24 = 323

The calculations above suggest that the chance of frost damage is only 2.5 % on October 2nd (day 275), 16 % on October 15th (day 287), 50 % on October 27th (day 299), 84 % on November 8th (day 311), and 97.5 % on November 20th (day 323).

(f) Hypothesis Testing

{PRIVATE**}**In **section *3b***, we discovered that an important component of scientific method was the testing of ***hypotheses*** either through ***experiments*** or ***predictive*** forms of analysis. A hypothesis can be defined as a tentative assumption that is made for the purpose of empirical scientific testing. A hypothesis becomes a theory of science when repeated testing produces the same conclusion.

In most cases, hypothesis testing involves the following structured sequence of steps. The first step is the formulation of a ***null hypothesis***. The null hypothesis is the assumption that will be maintained by the researcher unless the analysis of data provides significant evidence to disprove it. The null hypothesis is denoted symbolically as **H0**. For example, here is a formulated null hypothesis related to the investigation of precipitation patterns over adjacent rural and urban land-use types:

H0 : There is no difference in precipitation levels between urban and adjacent rural areas.

The second step of hypothesis testing is to state the ***alternative hypothesis*** (**H1**). Researchers should structure their tests so that all outcomes are anticipated before the tests and that results can be clearly interpreted. Some tests may require the formulation of multiple alternative hypotheses. However, interpretation is most clear cut when the hypothesis is set up with only one alternative outcome. For the example dealing with precipitation patterns over adjacent rural and urban land-use types, the alternative might be:

H1 : There is an increase in precipitation levels in urban areas relative to adjacent rural areas because of the heating differences of the two surface types (the urban area heats up more and has increased convective uplift).

Step three involves the collection of data for hypothesis testing. It is assumed that this data is gathered in an unbiased manner. For some forms of analysis that use ***inferential*** statistical tests the data must be collected randomly, data observations should be independent of each other, and the variables should be normally distributed.

The fourth step involves testing the null hypothesis through predictive analysis or via experiments. The results of the test are then interpreted (acceptance or rejection of the null hypothesis) and a decision may be made about future investigations to better understand the system under study. In the example used here, future investigations may involve trying to determine the mechanism responsible for differences in precipitation between rural and urban land-use types.

Inferential Statistics and Significance Levels

Statisticians have developed a number of mathematical procedures for the purpose of testing hypotheses. This group of techniques is commonly known as ***inferential statistics*** (see **sections *3g*** and ***3h*** for examples). Inferential statistics are available both for predictive and experimental hypothesis testing. This group of statistical procedures allow researchers to test assumptions about collected data based on the laws of probability. Tests are carried out by comparing calculated values of the test statistic to assigned critical values.

For a given null hypothesis, the calculated value of the test statistic is compared with tables of critical values at specified significance levels based on probability. For example, if a calculated test statistic exceeds the critical value for a significance level of 0.05 then this means that values of the test statistic as large as, or larger than calculated from the data would occur by chance less than 5 times in 100 if the null hypothesis was indeed correct. In other words, if we were to reject the null hypothesis based on this probability value of the test statistic, we would run a risk of less than 5 % of acting falsely.

One-tailed and Two-tailed Tests

When using some types of inferential statistics the alternative hypothesis may be directional or non-directional. A directional hypothesis (or one-sided hypothesis) is used when either only positive or negative differences are of interest in an experimental study. For example, when an alternative hypothesis predicts that the mean of one sample would be greater (but not less) than another, then a directional alternative would be used. This type of statistical procedure is known as a ***one-tailed test***. A non-directional (or two-sided) hypothesis would be used when both positive and negative differences are of equal importance in providing evidence with which to test the null hypothesis. We call this type of test ***two-tailed***.

(g) Inferential Statistics: Comparison of Sample Means

{PRIVATE}Introduction

In Physical Geography it is often necessary to test whether two samples of the same natural event are statistically distinct under different conditions. For example, a research scientist may want to determine if rainfall from convective storms differs in its intensity in rural and adjacent urban landscapes. Information from this investigation can then be used to test theories concerning the influence of the urban landscape on the development of thunderstorms. Further, this type of ***hypothesis testing*** changes our general comprehension in Physical Geography from simple description to process oriented understanding. A number of inferential statistical procedures have been developed to carry out this process. We will examine two of the most popular statistical techniques available to researchers.

Mann-Whitney U test

In research work it is often necessary to test whether two samples of the same phenomenon are statistically different. One test that is particularly useful for this type of test situation is the **Mann-Whitney U test**. This technique is ***non-parametric*** or '**distribution-free**' in nature. Non-parametric methods are particularly suited to data that are not ***normally distributed***.

Setting up the Null Hypothesis

This is the first stage of any statistical analysis and states the hypothesis that is to be tested. This is the assumption that will be maintained unless the data provide significant evidence to discredit it. The null hypothesis is denoted symbolically as **H0**. For our example, the ***null hypothesis*** would be:

H0 : there is no difference in precipitation levels between urban and adjacent rural areas.

It is also necessary to state the ***alternative hypothesis*** (**H1**). In this case the alternative might be:

H1 : there is an increase in precipitation levels in urban areas relative to adjacent rural areas because of the heating differences of the two surface types (the urban area heats up more and has increased convective uplift).

Calculation

To calculate the **U-statistic**, the values for both sets of samples are ranked together in an ascending fashion. When ties occur, the mean rank of all the scores involved in the tie is entered for those observations. The rank values for each set of observations are then summed separately to determine the following values:

S **r1** and S **r2**

These values are then entered in the formulae shown under **Table 3g-1** for the calculation of **U** and **U1**.

{PRIVATE**}Table 3g-1:** Analysis of convective precipitation levels per storm event (mm of rain) between urban and rural areas using the Mann-Whitney U test.

{PRIVATE**}Urban (n1)**	**Rural (n2)**	**Rank (n1)**	**Rank (n2)**
28	14	26	5
27	20	25	13.5
33	16	28	8.5
23	13	20	2.5
24	18	23	11
17	21	10	16
25	23	24	20
23	20	20	13.5
31	14	27	5
23	20	20	13.5
23	20	20	13.5
22	14	17	5
15	11	7	1
-	16	-	8.5
-	13	-	2.5
n1 = 13	**n2** = 15	S **r1** = 267	S **r2** = 139

Introduction

In Physical Geography it is often necessary to test whether two samples of the same natural event are statistically distinct under different conditions. For example, a research scientist may want to determine if rainfall from convective storms differs in its intensity in rural and adjacent urban landscapes. Information from this investigation can then be used to test theories concerning the influence of the urban landscape on the development of thunderstorms. Further, this type of ***hypothesis testing*** changes our general comprehension in Physical Geography from simple description to process oriented understanding. A number of inferential statistical procedures have been developed to carry out this process. We will examine two of the most popular statistical techniques available to researchers.

Mann-Whitney U test

In research work it is often necessary to test whether two samples of the same phenomenon are statistically different. One test that is particularly useful for this type of test situation is the **Mann-Whitney U test**. This technique is ***non-parametric*** or '**distribution-free**' in nature. Non-parametric methods are particularly suited to data that are not ***normally distributed***.

Setting up the Null Hypothesis

This is the first stage of any statistical analysis and states the hypothesis that is to be tested. This is the assumption that will be maintained unless the data provide significant evidence to discredit it. The null hypothesis is denoted symbolically as **H0**. For our example, the ***null hypothesis*** would be:

H0 : there is no difference in precipitation levels between urban and adjacent rural areas.

It is also necessary to state the ***alternative hypothesis*** (**H1**). In this case the alternative might be:

H1 : there is an increase in precipitation levels in urban areas relative to adjacent rural areas because of the heating differences of the two surface types (the urban area heats up more and has increased convective uplift).

Calculation

To calculate the **U-statistic**, the values for both sets of samples are ranked together in an ascending fashion. When ties occur, the mean rank of all the scores involved in the tie is entered for those observations. The rank values for each set of observations are then summed separately to determine the following values:

S **r1** and S **r2**

These values are then entered in the formulae shown under **Table 3g-1** for the calculation of **U** and **U1**.

{PRIVATE**}Table 3g-1:** Analysis of convective precipitation levels per storm event (mm of rain) between urban and rural areas using the Mann-Whitney U test.

{PRIVATE **}Urban (n1)**	**Rural (n2)**	**Rank (n1)**	**Rank (n2)**
28	14	26	5
27	20	25	13.5
33	16	28	8.5
23	13	20	2.5
24	18	23	11
17	21	10	16
25	23	24	20
23	20	20	13.5
31	14	27	5
23	20	20	13.5
23	20	20	13.5
22	14	17	5
15	11	7	1
-	16	-	8.5
-	13	-	2.5
n1 = 13	**n2** = 15	S **r1** = 267	S **r2** = 139

U = n1 •n2 + {n1•(n1 + 1)}/2 - S r1

U = 13 • 15 + {13•(13 + 1)}/2 - 267 = 19

U1 = n1•n2 + {n2•(n2 + 1)}/2 - S r2

U1 = 13 • 15 + {15•(15 + 1)}/2 - 139 = 176

where **n1** is the number of observations in the first sample, and **n2** is the number of observations in the second sample.

The lower of these two values (**U** and **U1**) is then taken to determine the significance of the difference between the two data sets. Calculated from the data found on **Table 3g-1**, the value of **U** is 19 and **U1** is 176. The lower value is thus 19. This value is now compared to the critical value found on the significance tables for the Mann-Whitney U (**Table 3g-2**) at a pre-determined significance level for the given sample sizes. An important feature of this statistical test is that the greater the difference between the two sets of samples, the smaller will be the test statistic (i.e., the lower value of **U** or **U1**). Thus, if the computed value is lower than the critical value in **Table 3g-2**, the null hypothesis (**H0**) is rejected for the given significance level. If the computed value is greater than the critical value, we then accept the null hypothesis.

Using a significance level of 0.05 with sample sizes of **n1** = 13 and **n2** = 15, the critical value in the table for a two-tailed test is 54. **Note** that this is a two-tailed test, because the direction of the relationship is not specified. The computed value of U is 19, which is much less than the tabulated value. Thus, the null hypothesis (**H0**) is rejected and the alternative hypothesis (**H1**) is accepted.

Table 3g-2: Critical values of U for the Mann-Whitney U test (P = 0.05).

{PRIVATE} n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	-	-	-	-	-	-	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	-	0	1	2	2	3	5	6	7	8	9	10	12	13	14	15	17	18	19	20
6	-	-	-	-	-	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	-	-	-	-	-	-	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	-	-	-	-	-	-	-	13	15	17	19	22	24	26	29	31	34	36	38	41
9	-	-	-	-	-	-	-	-	17	20	23	26	28	31	34	37	39	42	45	48
10	-	-	-	-	-	-	-	-	-	23	26	29	33	36	39	42	45	48	52	55
11	-	-	-	-	-	-	-	-	-	-	30	33	37	40	44	47	51	55	58	62
12	-	-	-	-	-	-	-	-	-	-	-	37	41	45	49	53	57	61	65	69
13	-	-	-	-	-	-	-	-	-	-	-	-	45	50	54	59	63	67	72	76
14	-	-	-	-	-	-	-	-	-	-	-	-	-	55	59	64	67	74	78	83
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	64	70	75	80	85	90
16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	75	81	86	92	98
17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	87	93	99	105
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99	106	112
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	113	119
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	127

Unpaired Student's t-Test

Another statistical test used to determine differences between two samples of the same phenomenon is the **Student's t-test**. The Student's t -test, however, differs from the Mann-Whitney U test in that it is used with data that is ***normally distributed*** (***parametric***).

Table 3g-3 describes the data from two "treatments" of strawberry plants that were subjected to freezing temperatures over an equal period of days. The data displayed are the numbers of fruit produced per plant. The treatments consist of genetically engineered and control (normal) varieties.

H0 : there is no difference in the number of strawberries produced by the control and genetically engineered varieties.

H1 : there is a difference in the numb er of strawberries produced by the control and genetically engineered varieties.

Table 3g-3: Strawberry data.

{PRIVATE}**Control (Xa)**	**(Xa)^2**	**Engineered (Xb)**	**(Xb)^2**
10.7	114.49	10.0	100
6.7	44.89	10.2	104.04
8.7	75.69	12.0	144
8.3	68.89	10.5	110.25
10.6	112.36	10.3	106.09
8.3	68.89	9.4	88.36
10.0	100	9.7	94.09
9.8	96.04	12.7	161.29
9.1	82.81	10.4	108.16
9.8	96.04	10.8	116.64
8.9	79.21	12.3	151.29
10.3	106.09	11.0	121
8.3	68.89	12.3	151.29
9.4	88.36	10.8	116.64
8.8	77.44	10.6	112.36
10.9	118.81	10.1	102.01
9.4	88.36	10.7	114.49
7.9	62.41	10.2	104.04
8.3	68.89	9.5	90.25
8.6	73.96	11.0	121
11.1	123.21	9.4	88.36
8.8	77.44	10.2	104.04
7.5	56.25	11.2	125.44
8.9	79.21	10.5	110.25
7.9	62.41	11.9	141.61
-	-	12.3	151.29
S **Xa** = 227	S **Xa^2** = 2091.04	S **Xb** = 280	S **Xb^2** = 3038.28
(S **Xa)^2** = 51,529		**(**S **Xb)^2** = 78,400	

$\bar{X}a$= 9.08

$\bar{X}b$= 10.77

na = 25 **nb** = 26

To test the hypothesis that there is no difference between strawberry varieties we compute:

$$t = \frac{\bar{X}a - \bar{X}b}{\sqrt{\frac{S^2(na + nb)}{(na)(nb)}}}$$

where : $\bar{X}a$ and $\bar{X}b$ are the arithmetic means for groups **A** and **B**, **na** and **nb** are the number of observations in groups **A** and **B**, and S^2 is the pooled within-group variance.

To compute the pooled within variance, we calculate the corrected sum of squares (**SS**) within each treatment group.

$$SSa = \Sigma Xa^2 - \frac{(\Sigma Xa)^2}{na} = 2091.04 - \frac{51529}{25}$$

= 2091.04 - 2061.16 = 29.88

$$SSb = \Sigma Xb^2 - \frac{(\Sigma Xb)^2}{nb} = 3038.28 - \frac{78400}{26}$$

= 3038.28 - 3015.38 = 22.90

Then the pooled variance is

$$S^2 = \frac{(SSa + SSb)}{(na - 1) + (nb - 1)} = \frac{52.78}{49}$$

= 1.077

and,

$$t = \frac{10.77 - 9.08}{\sqrt{\frac{1.077(25 + 26)}{(25)(26)}}} = \frac{1.69}{\sqrt{0.08450}}$$

= 5.814

This value of t has (**na** - 1) + (**nb** - 1) degrees of freedom. If it exceeds the tabular value of t (**Table 3g-4**) at a pre-determined probability level, we can reject the null hypothesis, and the difference between the two means would be considered statistically significant (greater than would be expected by chance if there is actually no difference).

In this case, the critical t value with **49 degrees of freedom** at the **0.01** probability level is approximately **2.682**. Since our sample is greater than this, the difference is significant at the 0.01 level and we can reject the null hypothesis.

The Paired t-Test

The previous description for the t-test assumed that the random samples are drawn from the two populations independently. However, there are some situations where the observations are paired. Analyzing paired data is done differently than if the two samples are independent. This modified procedure is known as a **paired t-test**. Most statistical software programs that perform the Student's t-test have options to select for either a paired or unpaired analysis.

Table 3g-4: Critical values of Student's t-distribution (2-tailed).

{PRIVATE}**Degrees of Freedom**	**P=0.10**	**P=0.05**	**P=0.02**	**P=0.01**	**P=0.001**	**Degrees of Freedom**
1	6.314	12.706	31.821	63.657	636.619	**1**
2	2.920	4.303	6.965	9.925	31.598	**2**

3	2.353	3.182	4.541	5.841	12.924	**3**
4	2.132	2.776	3.747	4.604	8.610	**4**
5	2.015	2.571	3.365	4.032	6.869	**5**
6	1.943	2.447	3.143	3.707	5.959	**6**
7	1.895	2.365	2.998	3.499	5.408	**7**
8	1.860	2.306	2.896	3.355	5.041	**8**
9	1.833	2.262	2.821	3.250	4.781	**9**
10	1.812	2.228	2.764	3.169	4.587	**10**
11	1.796	2.201	2.718	3.106	4.437	**11**
12	1.782	2.179	2.681	3.055	4.318	**12**
13	1.771	2.160	2.650	3.012	4.221	**13**
14	1.761	2.145	2.624	2.977	4.140	**14**
15	1.753	2.131	2.602	2.947	4.073	**15**
16	1.746	2.120	2.583	2.921	4.015	**16**
17	1.740	2.110	2.567	2.898	3.965	**17**
18	1.734	2.101	2.552	2.878	3.922	**18**
19	1.729	2.093	2.539	2.861	3.883	**19**
20	1.725	2.086	2.528	2.845	3.850	**20**
21	1.721	2.080	2.518	2.831	3.819	**21**
22	1.717	2.074	2.508	2.819	3.792	**22**
23	1.714	2.069	2.500	2.807	3.767	**23**
24	1.711	2.064	2.492	2.797	3.745	**24**
25	1.708	2.060	2.485	2.787	3.725	**25**
26	1.706	2.056	2.479	2.779	3.707	**26**
27	1.703	2.052	2.473	2.771	3.690	**27**
28	1.701	2.048	2.467	2.763	3.674	**28**
29	1.699	2.045	2.462	2.756	3.659	**29**
30	1.697	2.042	2.457	2.750	3.646	**30**
40	1.684	2.021	2.423	2.704	3.551	**40**
60	1.671	2.000	2.390	2.660	3.460	**60**
120	1.658	1.980	2.358	2.617	3.373	**120**

(h) Inferential Statistics: Regression and Correlation

{PRIVATE}Introduction

Regression and **correlation analysis** are statistical techniques used extensively in physical geography to examine causal relationships between variables. Regression and correlation measure the degree of **relationship** between two or more variables in two different but related ways. In regression analysis, a single ***dependent*** variable, **Y**, is considered to be a function of one or more independent variables, **X**1, **X**2, and so on. The values of both the ***dependent*** and ***independent*** variables are assumed as being ascertained in an error-free random manner. Further, parametric forms of regression analysis assume that for any given value of the independent variable, values of the dependent variable are normally distributed about some mean. Application of this statistical procedure to dependent and independent variables produces an equation that "best" approximates the functional relationship between the data observations.

Correlation analysis measures the degree of association between two or more variables. ***Parametric*** methods of correlation analysis assume that for any pair or set of values taken under a given set of conditions, variation in each of the variables is random and follows a normal distribution pattern. Utilization of correlation analysis on dependent and independent variables produces a statistic called the ***correlation coefficient*** (**r**). The square of this statistical parameter (the ***coefficient of determination*** or $\mathbf{r^2}$) describes what proportion of the variation in the dependent variable is associated with the **regression** of an independent variable.

Analysis of variance is used to test the significance of the **variation** in the dependent variable that can be attributed to the **regression** of one or more independent variables. Employment of this statitical procedure produces a **calculated F-value** that is compared to a **critical Fvalues** for a particular level of statistical probability. Obtaining a significant calculated F-value indicates that the results of regression and correlation are indeed true and not the consequence of chance.

Simple Linear Regression

In a simple regression analysis, one ***dependent variable*** is examined in relation to only one ***independent variable***. The analysis is designed to derive an equation for the line that best models the relationship between the dependent and independent variables. This equation has the mathematical form:

$$\mathbf{Y = a + bX}$$

where, **Y** is the value of the dependent variable, **X** is the value of the independent variable, **a** is the intercept of the regression line on the **Y** axis when **X** = 0, and **b** is the slope of the regression line.

The following **table** contains randomly collected data on growing season precipitation and cucumber yield (**Table 3h-1**). It is reasonable to suggest that the amount of water received on a field during the growing season will influence the yield of cucumbers growing on it. We can use this data to illustate how regression analysis is carried out. In this table, precipitation is our independent variable and is not affected by variation in cucumber yield. However, cucumber yield is influenced by precipitation, and is therefore designated as the Y variable in the analysis.

{PRIVATE}**Precipitation mm (X)**	**Cucumbers kilograms per m^2 (Y)**	**Precipitation mm (X)**	**Cucumbers kilograms per m^2 (Y)**
22	.36	103	.74
6	.09	43	.64
93	.67	22	.50
62	.44	75	.39
84	.72	29	.30
14	.24	76	.61
52	.33	20	.29
69	.61	29	.38
104	.66	50	.53
100	.80	59	.58
41	.47	70	.62
85	.60	81	.66
90	.51	93	.69
27	.14	99	.71
18	.32	14	.14
48	.21	51	.41
37	.54	75	.66
67	.70	6	.18
56	.67	20	.21
31	.42	36	.29
17	.39	50	.56
7	.25	9	.13

2	.06	2	.10
53	.47	21	.18
70	.55	17	.17
6	.07	87	.63
90	.69	97	.66
46	.42	33	.18
36	.39	20	.06
14	.09	96	.58
60	.54	61	.42

S **X = 3,050**

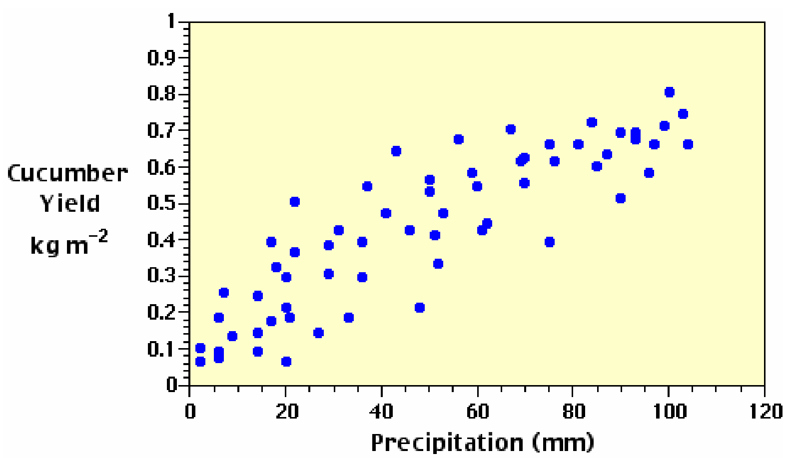
$\bar{X}$ **= 49.1935**

S **Y = 26.62**

$\bar{Y}$ **= 0.4294**

n = 62

Often the first step in regression analysis is to plot the **X** and **Y** data on a graph (**Figure 3h-1**). This is done to graphically visualize the relationship between the two variables. If there is a simple relationship, the plotted points will have a tendancy to form a recognizable pattern (a straight line or curve). If the relationship is strong, the pattern will be very obvious. If the relationship is weak, the points will be more spread out and the pattern less distinct. If the points appear to fall pretty much at random, there may be no relationship between the two variables.



{PRIVATE}**Figure 3h-1**: Scattergram plot of the precipitation and cucumber yield data found in **Table 3h-1**. The distribution of the data points indicates a possible positive linear relationship between the two variables.

The type of pattern (straight line, parabolic curve, exponential curve, etc.) will determine the type of regression model to be applied to the data. In this particular case, we will examine data that produces a simple straight-line relationship (see **Figure 3h-1**). After selecting the model to be used, the next step is to calculate the corrected sums of squares and products used in a bivariate linear regression analysis. In the following equations, capital letters indicate uncorrected values of the variables and lower-case letters are used for the corrected parameters in the analysis.

The **corrected sum of squares** for **Y**:

$$S\ y^2 = S\ Y^2 - \frac{(\Sigma Y)^2}{n}$$

= $(0.36^2 + 0.09^2 + ... + 0.42^2) - (26.62^2) / 62$

= 2.7826

The **corrected sum of squares** for **X**:

$$S\ x^2 = S\ x^2 - \frac{(\Sigma X)^2}{n}$$

= $(22^2 + 6^2 + ... + 61^2) - (3,050^2) / 62$

= 59,397.6775

The **corrected sum of products**:

$$S\ xy = S\ (XY) - \frac{(\Sigma X)(\Sigma Y)}{n}$$

= ((22)(.36) + (6)(.09) + ... + (61)(.42)) - ((26.62)(3,050)) / 62

= 354.1477

As discussed earlier, the general form of the equation for a straight line is **Y = a + bX**. In this equation, **a** and **b** are constants or regression coefficients that are estimated from the data set. Based on the mathematical procedure of least squares, the best estimates of these coefficients are:

$$b = \frac{(\Sigma xy)}{(\Sigma x^2)}$$

= (354.1477) / (59,397.6775) = 0.0060

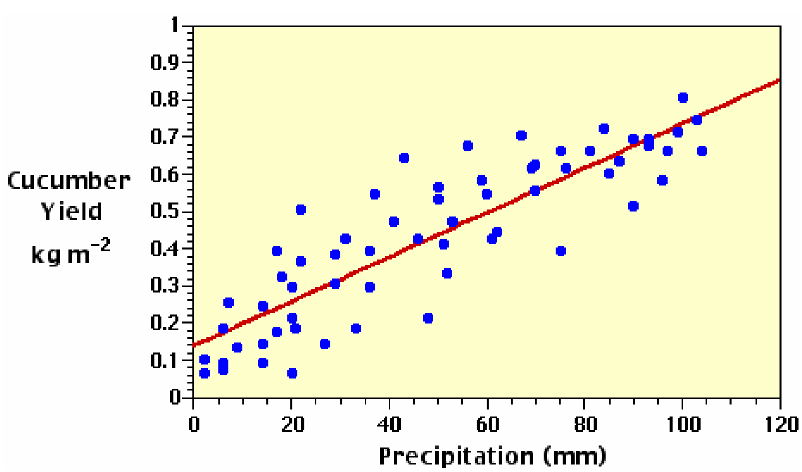
a = Y - bX = 0.42935 - (0.0060)(49.1935) = 0.1361

Substituting these estimates into the general linear equation suggests the following relationship between the **Y** and **X** variables:

$\hat{Y}$= 0.1361 + 0.0060X

where $\hat{Y}$ indicates that we are using an estimated value of **Y**.

With this equation, we can estimate the the number of cucumbers (**Y**) from the measurements of precipitation (**X**) and describe this relationship on our **scattergram** with a best fit straight-line (**Figure 3h-2**). Because **Y** is estimated from a known value of **X**, it is called the ***dependent variable*** and **X** the ***independent variable***. In plotting the data in a graph, the values of **Y** are normally plotted along the vertical axis and the values of **X** along the horizontal axis.



{PRIVATE}**Figure 3h-2**: Scattergram plot of the precipitation and cucumber yield data and the regression model best fit straight-line describing the linear relationship between the two variables.

Regression Analysis and ANOVA

A regression model can be viewed of as a type of moving average. The regression equation attempts to explain the relationship between the Y and X variables through linear association. For a particular value of **X**, the regression model provides us with an estimated value of **Y**. Yet, **Figure 3h-2** indicates that many of the plotted values of the actual data are observed to be above the regression line while other values are found below it. These variations are caused either by sampling error or the fact that some other unexplained independent variable influences the individual values of the **Y** variable.

The **corrected sum of squares** for **Y** (i.e., S **y**^{2}) determines the total amount of variation that occurs with the individual observations of **Y** about the mean estimate of $\overline{Y}$. The amount of variation in **Y** that is directly related with the regression on **X** is called the **regression sum of squares**. This value is calculated accordingly:

Regression SS = $\frac{(\Sigma xy)^2}{\Sigma x^2}$ **= (354.1477)^2 / (59,397.6775) = 2.1115**

As discussed above, the total variation in **Y** is determined by S **y^2 = 2.7826**. The amount of the total variation in **Y** that is not associated with the regression is termed the **residual sum of squares**. This statistical paramter is calculated by subtracting the **regression sum of squares** from the **corrected sum of squares** for **Y** (S **y**^{2}):

Residual SS = S **y**^{2} **- Regression SS**

= 2.7826 - 2.1115

= 0.6711

The unexplained variation can now be used as a standard for testing the amount of variation attributable to the regression. Its significance can be tested with the **F test** from calculations performed in an **Analysis of Variance** table.

{PRIVATE}**Source of variation**	**df 1**	**SS**	**MS 2**
Due to regression	1	2.1115	2.1115
Residual (unexplained)	60	0.6711	0.0112
Total	61	2.7826	-

1 There were 62 values of **Y** analyzed and therefore n = 62. The total sum of squares **degrees of freedom** (df) is determined as n-1 or 61. The regression of **Y** on **X** has 1 **degree of freedom**. The residual or unexplained **degrees of freedom** is determined by subtracting regression df (1) from total sum of squares df (61).

2 MS is calculated as **SS / df**.

Using the **Analysis of Variance** procedure, the regression is tested by determining the calculated **F statistic**:

F = (Regression MS) / (Residual SS) = (2.1115) / (0.0112) = 188.86

To test this statistic we use a **table of F** to determine a **critical test value** for a probability of 0.01 or 1 % (this relationship can occur by chance only in 1 out 100 cases) and with **1,60 degrees of freedom.** According to the table the **critical test value** is **7.1**. In this test, the relationship is deemed significant if the calculated **F statistic** is greater than the **critical test value**. This regression is statistically significant at the **0.01 level** because **188.86** is greater than **7.1**.

Caution must be taken when interpreting the results of regression. In our example, we found a significant relationship between precipitation and cucumber yield. However, this conclusion may not be the result of a causal relationship between the two variables. A third variable that is directly associated to both precipitation and cucumber yield may be **confounding** the interpretation of the analysis. Absolute verification of associations between variables can only be confirmed with experimental manipulation.

Coefficient of Determination

To measure how strong the correlation is between the two variables, we can determine the amount of the total variation in **Y** that is associated with the regression model. This ratio is sometimes called the ***coefficient of determination*** and is represented by the symbol $\mathbf{r^2}$. The value of the coefficient of determination ranges from 1.00 to 0.00. The calculated **coefficient of determination** from the data set above was 0.76 or 76 % (as calculated below). This value suggests that 76 % of the **variation** in **Y** was associated with the change seen **X** from the data set observations.

Coefficient of determination = (Regression SS) / (Total SS)

= (2.1115) / (2.7826) = 0.7588

Correlation Coefficient

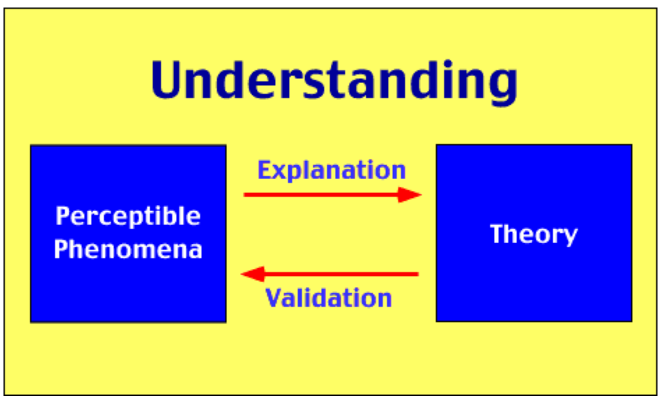
Another useful regression statistic that measures the strength of the correlation between to variables is the ***correlation coefficient*.**This statistic is often represented by the symbol "**r**" and is determined by taking the square-root of the **coefficient of determination** . The value of the correlation coefficient ranges from 1.00 to -1.00. A value of 0.0 indicates that there is absolutely no relationship between the **X** and **Y** variables. The strength of the relationship between the X and Y variables increases as the value of r approaches 1.00 and -1.00. Perfect correlation occurs if r equals either 1.00 (perfect positive) or -1.00 (perfect negative). Positive correlation coefficients indicate that an increase in the value of the **X** variable results in an increase in the value of the **Y** variable. Negative correlation coefficients indicate that an increase in the value of the **X** variable results in a decrease in the value of the **Y** variable.

4) Introduction to Systems Theory

(a) Humans and Their Models

{PRIVATE}The world of nature is very complex. In order to understand this complexity humans usually try to visualize the phenomena of nature as a ***system***. A system is a set of interrelated components working together towards some kind of process. One of the simplest forms of a system is a ***model***. Both models and systems are **simplified versions** of **reality**.

We can use the following **graphical model** (or system) to illustrate the process of scientific understanding. This process involves a continuous interaction between **perceptible phenomena** and ***theory***. This **model** is shown below:



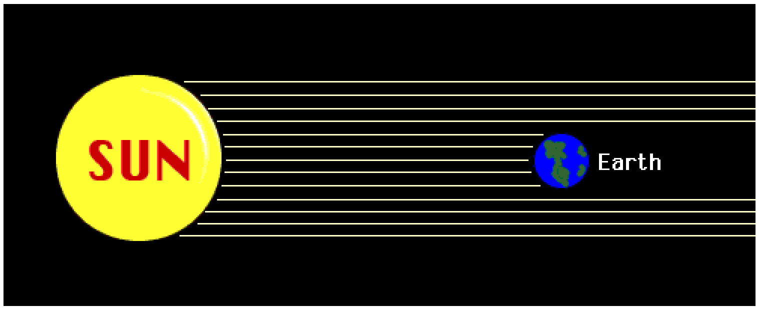
{PRIVATE}**Figure 4a-1:** Relationship between phenomena, theory, and understanding using scientific method. The interaction between perceptible phenomena and theory is accomplished through explanation and validation.

This simple model, while an extreme abstraction of reality, illustrates how scientific understanding works. It suggests that in scientific understanding **perceptible phenomena** and ***theory*** interact through **explanation** and **validation**.

In Physical Geography, and many other fields of knowledge, systems and models are used extensively as aids in explaining natural phenomena around us.

(b) Definitions of Systems and Models

{PRIVATE}As suggested in the previous section, a ***system*** is a group of parts that interact according to some kind of process (see **Figure 4b-1**). Systems are often visualized or ***modeled*** as **component blocks** with some kind of **connections** drawn. For example, the illustration below describes the interception of solar radiation by the Earth. In this system, the Earth and sun, the **parts** or **component blocks**, are represented by two colored circles of different size. The process of solar emission and the interception of the sun's radiation by the Earth (the **connection**) is illustrated by the drawn lines.



{PRIVATE**}Figure 4b-1:** Simple visual model of solar radiation being emitted from the sun and intercepted by the Earth.

All systems have the same **common characteristics**. These common characteristics are summarized below:

All systems have some **structure**.

All systems are **generalizations** of **reality**.

They all **function** in the same way.

There are **functional** as well as **structural relationships** between the units of a system.

Function implies the **flow** and **transfer** of some **material**. Systems exchange energy and matter internally and with their surrounding environment through various processes of ***input*** and ***output***.

Function requires the presence of some **driving force**, or some source of energy.

All systems show some degree of **integration**.

Within its defined ***boundary*** the system has three kinds of **properties**:

Elements - are the kinds of things or substances composing the system. They may be atoms or molecules, or larger bodies of matter-sand grains, rain drops, plants, or cows.

Attributes - are characteristics of the elements that may be perceived, for example: quantity, size, color, volume, temperature, and mass.

Relationships - are the associations that exist between elements and attributes based on cause and effect.

The ***state*** of the system is defined when each of its **properties** (e.g. ***elements***, ***attributes***, and ***relationships***) has a defined **value**.

Scientists have examined and **classified** many types of systems. These types include:

Isolated System - a system where there are no interactions outside its boundary layer. Such systems are common in laboratory experiments.

Closed System - is closed with respect to matter, but energy may be transferred between the system and its surroundings. Earth is essentially a closed system.

Open System - is a system where both matter and energy can cross the boundary of the system. Most environmental systems are open.

Morphological System - is a system where we understand process relationships or correlations between the elements of the system in terms of measured features.

Cascading System - in these systems we are interested in the movement of energy and/or matter from one element to another and understand the processes involved.

Process-Response System - in these systems we are interested in the movement, storage and transformation of energy and matter and the relationships between measured features in the various elements of the system.

Control System - a system that is intelligently manipulated by humans.

Ecosystem - is concerned with the biological relationships within the environment and the interactions between organis ms and their physical surroundings.

(c) Structure of Systems

{PRIVATE**}**Systems exist at every scale of size and are often arranged in some kind of **hierarchical fashion**. Large systems are often composed of one or more smaller systems working within its various ***elements***. Processes within these smaller systems can often be connected directly or indirectly to processes found in the larger system. A good example of a **system within systems** is the **hierarchy** of systems found in our ***Universe***. Let us examine this system from top to bottom:

At the highest level in this hierarchy we have the system that we call the cosmos or Universe. The elements of this system consist of galaxies, quasars, black holes, stars, planets and other heavenly bodies. The current structure of this system is thought to have come about because of a massive explosion known as the ***Big Bang*** and is controlled by gravity, weak and strong atomic forces, and electromagnetic forces.

Around some stars in the universe we have an obvious arrangement of planets, asteroids, comets and other material. We call these systems ***solar systems***. The elements of this system behave according to set laws of nature and are often found orbiting around a central star because of gravitational attraction. On some planets conditions may exist for the development of dynamic interactions between the hydrosphere, lithosphere, atmosphere, or biosphere.

We can define a **planetary system** as a celestial body in space that orbits a star and that maintains some level of dynamics between its lithosphere, atmo sphere and hydrosphere. Some planetary systems, like the Earth, can also have a biosphere. If a planetary system contains a biosphere, dynamic interactions will develop between this system and the lithosphere, atmosphere and hydrosphere. These interactions can be called an ***environmental system***. Environmental systems can also exist at smaller scales of size (e.g., a single flower growing in a field could be an example of a small-scale environmental system).

The Earth's biosphere is made up small interacting entities called ***ecosystems***. In an ecosystem, populations of species group together into ***communities*** and interact with each other and the ***abiotic*** environment. The smallest lining entity in an ecosystem is a single organism. An organism is alive and functioning because it is a **biological system**. The elements of a biological system consist of ***cells*** and larger structures known as ***organs*** that work together to produce life. The functioning of cells in any biological system is dependent on numerous chemical reactions. Together these chemical reactions make up a **chemical system**. The types of chemical interactions found in chemical systems are dependent on the atomic structure of the reacting matter. The components of atomic structure can be described as an **atomic system**.

(d) Environmental Systems as Energy Systems

{PRIVATE**}**In the previous section we define an ***environmental system*** as a system where life interacts with the various ***abiotic*** components found in the atmosphere, hydrosphere, and lithosphere. Environmental systems also involve the capture, movement, storage, and use of ***energy***. Thus, environmental systems are also **energy systems**.

In environmental systems, energy moves from the abiotic environment to life through processes like plant ***photosynthesis***. Photosynthesis packages this energy into simple organic compounds like ***glucose*** and ***starch***. Both of these organic molecules can be stored for future use. The following chemical formula describes how plants capture the sun's light energy and convert it into chemical energy:

$$\mathbf{6CO_2 + 6H_2O + light\ energy = C_6H_{12}O_6 + 6O_2}$$

The energy of light is used by plants in this reaction to chemically change carbon dioxide ($\mathbf{CO_2}$) and water ($\mathbf{H_2O}$) into oxygen ($\mathbf{O_2}$) and the energy rich organic molecule glucose ($\mathbf{C_6H_{12}O_6}$).

The chemical energy of photosynthesis can be passed on to other living or ***biotic*** components of an environmental system through ***biomass*** consumption or ***decomposition*** by ***consumer*** organisms. When needed for ***metabolic*** processes, the ***fixed*** organic energy stored in an organism can be released to do work via ***respiration*** or ***fermentation.***

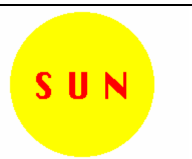
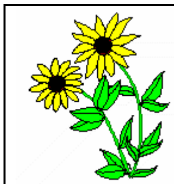
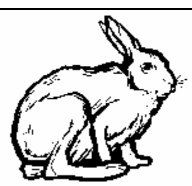
Energy also fuels a number of environmental processes that are essentially abiotic. For example, the movement of air by ***wind***, the ***weathering*** of ***rock*** into ***soil***, the formation of ***precipitation***, and the creation of mountains by ***tectonic*** forces. The first three processes derive their energy directly or indirectly from the sun's radiation that is received at the Earth's surface. Mountain building is fueled by the heat energy that exists within the Earth's interior.

Finally, the movement of energy in environmental systems always obeys specific ***thermodynamic laws*** that cannot be broken. We will learn more about these laws later in this textbook.

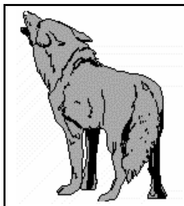
(e) Food Chain as an Example of a System

{PRIVATE**}**A ***food chain*** models the movement of ***energy*** in an ***ecosystem*** (a form of environmental system). **Figure 4e-1** below illustrates the movement of energy in a typical food chain. In this diagram, we begin the food chain with 100,000 units of light energy from the ***sun***. Note, the amount of energy available at each successive level (called ***trophic levels***) of this system becomes progressively less. Only 10 units of energy are available at the last level (***carnivores***) of the food chain. A number of factors limit the ***assimilation*** of energy from one level to the next. We will examine these factors later in the textbook.

{PRIVATE**}**

	The sun is the original source of energy, in the form of light, for the food chain. (100,000 Units of Energy)
	Plants capture approximately 1% of the available light energy from the sun for biomass production by way of photosynthesis. Photosynthesis can be described chemically as: Light Energy + 6C02 + 6H20 = C6H1206 + 602 (1,000 Units of Energy)
	Herbivores consume approximately 10% of the plant biomass produced in a typical food chain. (100 Units of Energy)

{PRIVATE}

	Carnivores capture and consume about 10% of the energy stored by the herbivores. (10 Units of Energy)

{PRIVATE}**Figure 4e -1:** Model of the grazing food chain showing the movement of energy through an ecosystem.

Why is the above illustration an example of a ***system***? The concept of what makes something a system was fully explained in **section *4b***. In this topic, it was suggested that all systems share the following seven common characteristics:

All systems have some **structure**. In the above example, the structure consists of the system's three types of properties. This system has the following ***elements***: the sun, plants, ***herbivores*** and ***carnivores***. Within this system the main characteristic, or ***attribute***, of the elements being perceived is units of ***energy***. The last component that makes up the structure of this system is the cause and effect relationships between the elements and attributes. For example, the sun creates energy via nuclear fusion. This energy is ***radiated*** from the sun's surface and received by the surface of the Earth. On the surface of the Earth plants capture some of this solar radiation in the chloroplasts that exist in their tissues. Through ***photosynthesis*** the plants convert the radiant energy into energy rich organic matter. Some of the energy fixed by the plants is passed on to herbivores through consumption. Finally, a portion of the energy ***assimilated*** by the herbivores is then passed on to carnivores through consumption.

All systems are **generalizations** of **reality**. The food chain process described above is a simple abstraction of what actually happens in a variety of different types of terrestrial ecosystems of much greater complexity.

They all **function** in the same way. All systems consist of groups of parts that interact according to some kind of process. In the food chain model, the parts are the ***sun***, ***plants***, ***herbivores*** and ***carnivores***. There are two main processes taking place in this system. The first involves the movement of energy, in the form of radiation, from the sun to the plants. The second process involves the movement of energy, in the form of organic molecules, from plants to herbivores, and then finally to carnivores through biomass consumption.

There are **functional** as well as **structural relationships** between the units of a system. The structure within the food chain is defined by the functional relationships between the elements and attributes of the system.

Function implies the **flow** and **transfer** of some **material**. Systems exchange energy and matter internally and with their surrounding environment through various processes of ***input*** and ***output***. The main material being transferred into this system (input) is energy in the form of ***solar radiation.*** The solar radiation is then fixed into ***organic matter*** (output) by way of ***photosynthesis*** in the ***plants***. ***Herbivores*** consume the constructed plant organic molecules for nutrition to run their metabolism. The herbivores then provide food for the ***carnivores***.

Function requires the presence of some **driving force**, or some source of energy. The driving force in the food chain is the sun.

All systems show some degree of **integration**. Integration in the food chain comes primarily from the process of evolution. It was through evolution that plants, herbivores, and carnivores came about and developed ecological associations between each other.

(f) Equilibrium Concepts and Feedbacks

{PRIVATE}**Equilibrium**

Equilibrium describes the average condition of a system, as measured through one of its elements or attributes, over a specific period of time. For the purposes of this online textbook, there are six types of equilibrium:

Steady state equilibrium is an average condition of a system where the **trajectory** remains unchanged in time.

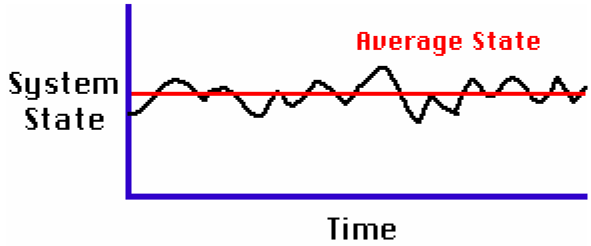


Figure 4f-1: Example of the state of a steady st ate equilibrium over time.

Thermodynamic equilibrium describes a condition in a system where the distribution of mass and energy moves towards maximum ***entropy***.

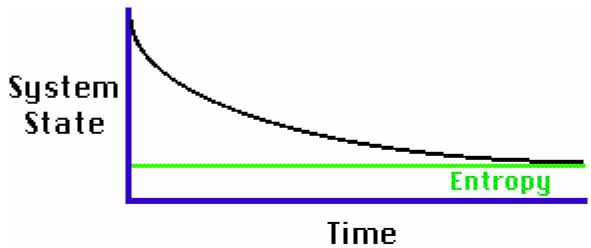


Figure 4f-2: Example of the state of a thermodynamic equilibrium over time.

(3) A ***dynamic equilibrium*** occurs when there are unrepeated average states through time.

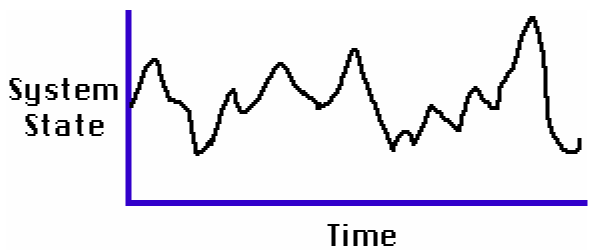


Figure 4f-3: Example of the state of a dynamic equilibrium over time.

(4) ***Static equilibrium*** occurs where force and reaction are balanced and the properties of the system remain unchanged over time.

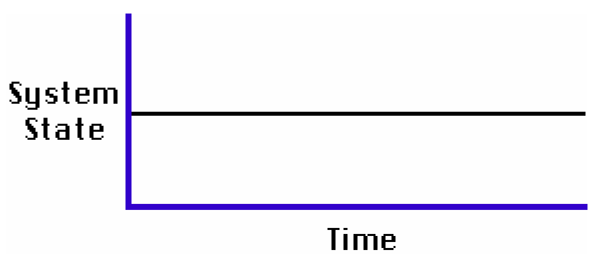


Figure 4f-4: Example of the state of a static equilibrium over time.

(5) In a ***stable equilibrium*** the system displays tendencies to return to the same equilibrium after ***disturbance***.

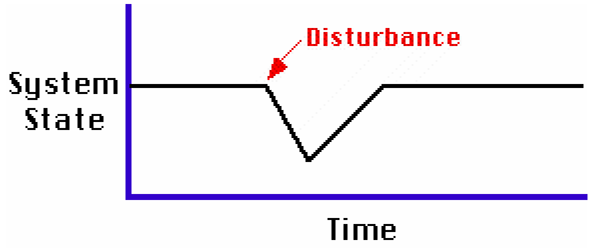


Figure 4f-5: Example of the state of a stable equilibrium over time.

(6) In an ***unstable equilibrium*** the system returns to a new equilibrium after ***disturbance***.

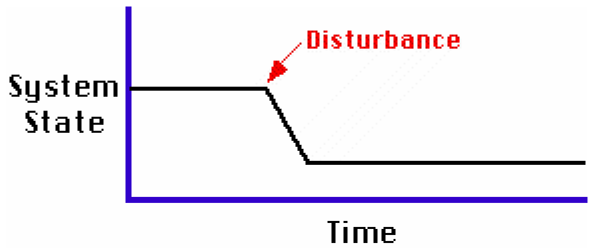


Figure 4f-6: Example of the state of an unstable equilibrium over time.

Feedbacks

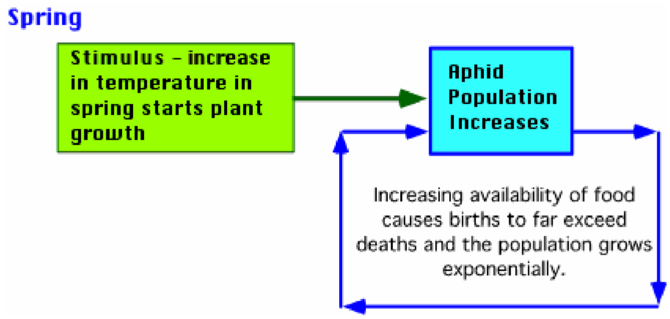
In order for a system to maintain a ***steady state*** or average condition the system must possess the capacity for ***self-regulation***. Self-regulation in many systems is controlled by ***negative feedback*** and ***positive feedback*** mechanisms. Negative-feedback mechanisms control the state of the system by dampening or reducing the size of the system's elements or attributes. Positive-feedback mechanisms feed or increase the size of one or more of the system's elements or attributes over time.

Interactions among living organisms, or between organisms and the abiotic environment typically involve both ***positive feedback*** and ***negative feedback*** responses. Feedback occurs when an organism's ***system state*** depends not only on some original stimulus but also on the results of its previous system state. Feedback can also involve the system state of non-living components in an ecosystem. A positive feedback causes a self-sustained change that increases the state of a system. Negative feedback causes the system to decease its state over time. The presence of both negative and positive feedback mechanisms in a system results in self-regulation.

To illustrate how these mechanisms work we can hypothetically examine the changes in aphid population growth in a mid-latitude climate in the following graphical models.

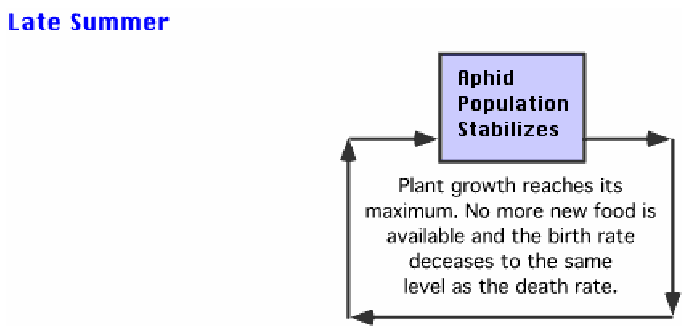
{PRIVATE}**Figure 4f-7:** Group of aphids feeding on the stem of a plant.

Over the winter only a few aphids survive the cold winter temperatures and scarcity of food (**Figure 4f-8**). However, warmer temperatures in the spring cause plants to start growing, providing the aphids with food. The continually increasing abundance of food increases the fertility of individual aphids and the aphid population starts to expand exponentially. This situation is a positive feedback as abundant food resources cause increased reproduction and rapid population growth.



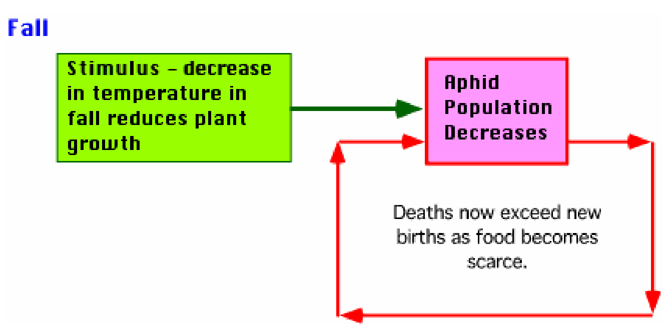
{PRIVATE}**Figure 4f-8:** Increase in food abundance causes a positive effect on the size of the aphid population. Births increase significantly and are much higher than deaths causing the population to expand.

By late summer the supply of plants has reached its maximum (**Figure 4f-9**). As a result, the availability of food to the growing aphid population becomes less per individual as time proceeds. Less food means lower egg production and births begin to decline. When births reach the same level as deaths the aphid population stops growing and stabilizes. The population is becoming too large for the food supply. The size of the population relative to food supply produces a negative feedback as the fertility of aphids begins to decline.



{PRIVATE}**Figure 4f-9:** As the abundance of food levels off aphid reproduction slows down and deaths begin to increase. Population size of the aphids begins to level off.

With the arrival of fall cooler temperatures reduce the supply of plants causing the large aphid population to be faced with further food scarcity (**Figure 4f-10**). The lack of food causes the aphid death rate to increase and the birth rate to decrease. The population begins to decline rapidly. The reduction of food caused by the cooler temperatures enhances the negative feedback that began in late summer.

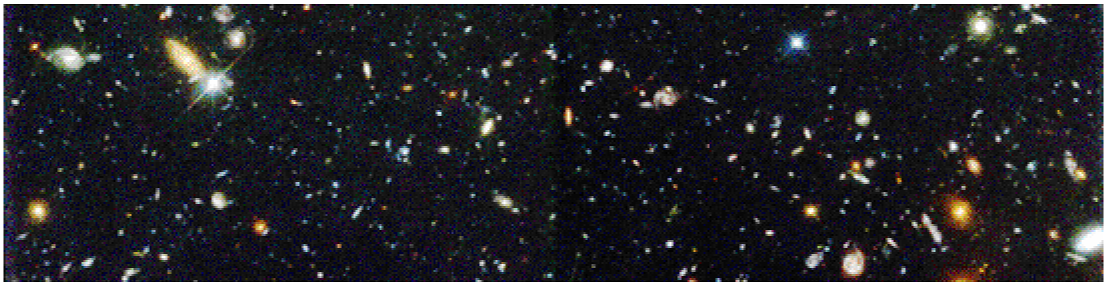


{PRIVATE}**Figure 4f-10:** The decrease in food abundance causes a negative effect on aphid population size. Deaths now exceed births and the population size of the aphids begins to decrease.

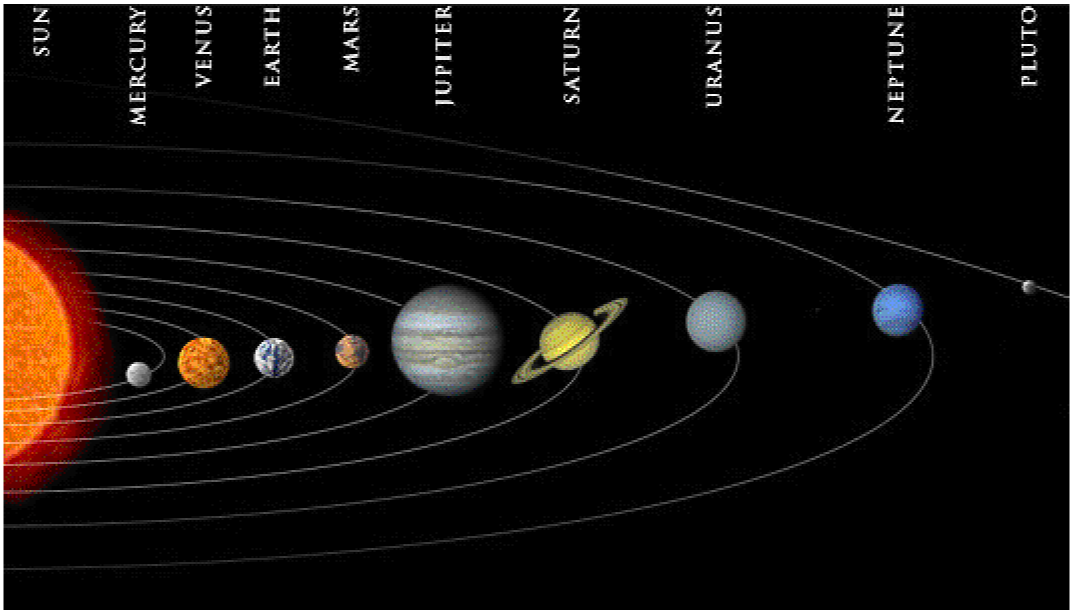
5) The Universe, Earth, Natural Spheres, and Gaia

(a) Evolution of the Universe

{PRIVATE}About 10 to 20 billion years ago all of the matter and energy in the ***Universe*** was concentrated into an area the size of an atom. At this instant, matter, energy, space and time did not exist. Then suddenly, the Universe began to expand at an incredible rate and matter, energy, space and time came into being (the ***Big Bang***). As the Universe expanded, matter began to coalesce into gas clouds, and then stars and planets. Our solar system formed about 5 billion years ago when the Universe was about 65 % of its present size (**Figure 5a-2**). Today, the Universe continues to expand.



{PRIVATE}**Figure 5a-1:** Hubble Space Telescope view of a distant cluster of galaxies near the beginning of time. (**Source:** *AURA/STScI*).



{PRIVATE}**Figure 5a-2:** Our solar system began forming about 5 billion years ago as gas clouds coalesce into planets and a star. Today, the solar system contains nine commonly recognized planets and the sun. (**Source:** *NASA*).

Why do Most Scientists Accept the Big Bang Theory?

The acceptance of this theory by the scientific community is based on a number of observations. These observations confirm specific predictions of the **Big Bang** theory. In a previous section, we learned that scientists test their theories through ***deduction*** and ***falsification***. Predictions associated with the **Big Bang** theory that have been tested by this process are:

If the Big Bang did occur, all of the objects within the Universe should be moving away from each other. In 1929, Edwin Hubble documented that the galaxies in our Universe are indeed moving away from each other.

The Big Bang should have left an "afterglow" from the explosion. In the 1960s, scientists discovered the existence of **cosmic background radiation**, the so-called "afterglow" after the Big Bang explosion. Our most accurate measurements of this cosmic radiation came in November 1989, by the Cosmic Background Explorer (COBE) satellite. The measurements from this satellite tested an important prediction of the **Big Bang** theory. This prediction suggests that the initial explosion that gave birth to the Universe should have created radiation with a spectrum that follows a blackbody curve. The COBE measurements indicated that the spectrum of the cosmic radiation varied from a blackbody curve by only 1 %. This level of error is considered insignificant.

If the Universe began with a Big Bang, extreme temperatures should have caused 25 percent of the mass of the Universe to become helium. This is exactly what is observed.

Matter in the Universe should be distributed homogeneously. Astronomical observations from the ***Hubble Space Telescope*** do indicate that matter in the Universe generally has a homogeneous distribution.

How will the Universe End?

Cosmologists have postulated two endings to the Universe. If the Universe is infinite or has no edge, it should continue to expand forever. A Universe that is finite or closed is theorized to collapse when expansion stops because of gravity. The collapse of the Universe ends when all matter and energy is compressed into the high energy, high-density state from which it began. This scenario is of course called the ***Big Crunch***. Some theorists have suggested that the Big Crunch will produce a new Big Bang and the process of an expanding Universe will begin again. This idea is called the **oscillating Universe theory**.

(b) Early History of the Earth

{PRIVATE**}**Scientists believe the Earth began its life about **4.6 billion** years ago. The Earth formed as cosmic dust lumped together to form larger and larger particles until 150 million years had passed. At about 4.4 billion years, the young Earth had a mass similar to the mass it has today. The continents probably began forming about 4.2 billion years ago as the Earth continued to cool. The cooling also resulted in the release of gases from the lithosphere, much of which formed the Earth's early atmosphere. Most of the Earth's early atmosphere was created in the first one million years after solidification (4.4 billion years ago). Carbon dioxide, nitrogen, and water vapor dominated this early atmosphere. **Table 5b-1** below describes the three major stages of development of the atmosphere.

Table 5b-1: Evolution of the Earth's atmosphere.

{PRIVATE**}Name of Stage**	**Duration of Stage (Billions of Years Ago)**	**Main Constituents of the Atmosphere**	**Dominant Processes and Features**
Early Atmosphere	4.4 to 4.0	H2O, hydrogen cyanide (HCN), ammonia (NH3), methane (CH4), sulfur, iodine, bromine, chlorine, argon	Lighter gases like hydrogen and helium escaped to space. All water was held in the atmosphere as vapor because of high temperatures.
Secondary Atmosphere	4.0 to 3.3	At 4.0 billion H2O, CO2, and nitrogen (N) dominant. Cooling of the atmosphere causes precipitation and the development of the oceans. By 3.0 billion CO2, H2O, N2 dominant. O2 begins to accumulate.	Continued release of gases from the lithosphere. Water vapor clouds common in the lower atmosphere. Chemosynthetic bacteria appear on the Earth at 3.6 billion. Life begins to modify the atmosphere.

Living Atmosphere	3.3 to Present	N_2 - 78%, O_2 - 21%, Argon - 0.9%, CO_2 - 0.036%	Development, evolution and growth of life increases the quantity of oxygen in the atmosphere from <1% to 21%. 500 million years ago concentration of atmospheric oxygen levels off. Humans begin modifying the concentrations of some gases in the atmosphere beginning around the year 1700.

As the Earth continued to cool, the water vapor found in the atmosphere condensed to form the oceans and other fresh water bodies on the continents. Oxygen began accumulating in the atmosphere through ***photo-dissociation*** of 02 from water, and by way of ***photosynthesis*** (life). The emergence of living organisms was extremely important in the creation of atmospheric oxygen and ***ozone***. Without ozone, life could not exist on land because of harmful ***ultraviolet radiation***.

Most of the build up of oxygen in the atmosphere occurred between 2.1 and 1.5 billion years ago as a direct result of photosynthesis from ocean based plants like algae. At about 450 million years ago, there was enough oxygen in the atmosphere to allow for the development of a stratospheric ozone layer that was thick enough to keep terrestrial life protected from ultraviolet radiation. As a result, terrestrial life began its development and expansion at this time. **Table 5b-2** describes the timing of the evolutionary development of some of the Earth's dominant forms of life before and after 450 million years before present (**BP**).

Table 5b-2: Approximate origin time of the major plant and animal groups.

{PRIVATE}**Organism Group**	**Time of Origin**
Marine Invertebrates	570 Million Years Ago
Fish	505 Million Years Ago
Land Plants	438 Million Years Ago
Amphibians	408 Million Years Ago
Reptiles	320 Million Years Ago
Mammals	208 Million Years Ago
Flowering Plants (*Angiosperms*)	140 Million Years Ago

(c) The Natural Spheres

{PRIVATE}From the standpoint of Physical Geography, the Earth can be seen to be composed of four principal components:

Lithosphere - describes the solid ***inorganic*** portion of the Earth (composed of ***rocks***, ***minerals*** and ***elements***). It can be regarded as the outer surface and interior of the solid Earth. On the surface of the Earth, the lithosphere is composed of three main types of rocks:

Igneous - rocks formed by solidification of molten ***magma***.

Sedimentary - rocks formed by the alteration and compression of old rock debris or ***organic*** sediments.

Metamorphic - rocks formed by alteration of existing rocks by intense heat or pressure.

Atmosphere - is the vast gaseous envelope of air that surrounds the Earth. Its boundaries are not easily defined. The atmosphere contains a complex system of gases and suspended particles that behave in many ways like fluids. Many of its constituents are derived from the Earth by way of chemical and biochemical reactions.

Hydrosphere - describes the waters of the Earth (see the ***hydrologic cycle***). Water exists on the Earth in various stores, including the ***atmosphere***, ***oceans***, ***lakes***, ***rivers***, ***soils***, ***glaciers***, and ***groundwater***. Water moves from one store to another by way of: ***evaporation***, ***condensation***, ***runoff***, ***precipitation***, ***infiltration*** and ***groundwater flow***.

Biosphere - consists of all living things, plant and animal. This zone is characterized by life in profusion, diversity, and ingenious complexity. Cycling of matter in this sphere involves not only metabolic reactions in organisms, but also many ***abiotic*** chemical reactions.

All of these spheres are interrelated to each other by dynamic interactions, like ***biogeochemical cycling***, that move and exchange both matter and energy between the four components.

(d) The Gaia Hypothesis

{PRIVATE**}**In 1965, ***J.E. Lovelock*** published the first scientific paper suggesting the ***Gaia hypothesis***. The Gaia hypothesis states that the temperature and composition of the Earth's surface are actively controlled by life on the planet. It suggests that if changes in the gas composition, temperature or oxidation state of the Earth are caused by extraterrestial, biological, geological, or other disturbances, life responds to these changes by modifying the abiotic environment through growth and metabolism. In simplier terms, biological responses tend to **regulate** the state of the Earth's environment in their favor.

The evidence for Gaia is as follows:

If not continually replaced by ***biotic*** activities gases like ***methane*** and **hydrogen** would become non-existant in the atmosphere in a few decades.

Carbon dioxide (C02) in the Earth's atmosphere is far less abundant than chemistry alone would allow. If life was deleted carbon dioxide would become 30 times more abundant. Large quantities of carbon dioxide are currently locked up by living organisms.

The sun's energy output has increased by 30 % in the past 3.5 billion years. Yet, historical climate data indicates that the temperature of the Earth has only fluctuated by about 5 degrees Celsius from the current ***average global temperature*** of 15 degrees Celsius. Computer climate models suggest that a 30 % reduction in solar radiation would create a global average temperature of between -10 and -52 degrees Celsius all things being equal. These results indicate that levels of atmospheric carbon dioxide must have been much higher in the past when the sun was less powerful. Extra atmospheric carbon dioxide would have created a greater ***greenhouse effect*** and warmer temperatures. These results also indicate that some **mechanism** must have removed carbon dioxide from the atmosphere as the sun's output of radiation increased over the Earth's geologic history. This mechanism is the conversion of atmospheric carbon dioxide into ***fossilized organic matter*** (***natural gas***, ***oil***, ***coal***, ***limestone***, and ***peat***). In other words, Gaia!

This theory is **important** to Physical Geography and other Earth Sciences for the following reasons:

The Gaia theory suggests that the abiotic and biotic environment is made up of many complex interrelationships;

Many of these complex interrelationships are quite delicate and may be altered by human activity to a breaking point; and

The theory suggests that humans must learn to respect Gaia by reducing their intentional modification of the Earth's abiotic and biotic components.

6) Energy and Matter

(a) Characteristics of Energy and Matter

{PRIVATE**}Introduction**

Energy is defined simply by scientists as the capacity for doing work. ***Matter*** is the material (***atoms*** and ***molecules***) that constructs things on the Earth and in the Universe. Albert Einstein suggested early in this century that energy and matter are related to each other at the atomic level. Einstein theorized that it should be possible to convert matter into energy. From Einstein's theories, scientists were able to harness the energy of matter beginning in the 1940s through nuclear ***fission***. The most spectacular example of this process is a nuclear explosion from an atomic bomb. A more peaceful example of our use of this fact of nature is the production of electricity from controlled fission reactions in nuclear reactors.

Einstein also suggested that it should be possible to transform energy into matter. In 1998, researchers at Stanford University's Linear Accelerator Center successfully converted energy into matter. This feat was accomplished by using lasers and incredibly strong electromagnetic fields to change ordinary light into matter. The results of this experiment may allow for the development of a variety of technological gadgets. One such development could be matter/energy transporters or food replicators that are commonly seen in some of our favorite science fiction programs.

Energy and matter are also associated to each other at much larger scales of nature. Later on in this chapter, we will examine how solar radiation provides the energy to create the matter that makes up organisms. Organisms then use some of this matter to power their ***metabolism***.

Types of Energy

Energy comes in a variety forms. The simplest definition of the types of energy suggests that two forms exist: ***kinetic energy*** and ***potential energy***. Kinetic energy is the energy due to motion. A rock falling from a cliff, a bee in flight, wind blowing leaves of trees, and water following over a waterfall are all examples of kinetic energy. Potential energy is the energy stored by an object that can be potentially transformed into another form of energy. Water stored behind a dam, the chemical energy of the food we consume, and the gasoline that we putting in our cars are all examples of potential energy. Conversion of this energy occurs when the energy in food is used for metabolism and it has been digested, when the water in the dam flows through turbines to produce electricity from motion, and when the gasoline is used in a engine to produce motion from combustion.

Some other forms of energy include heat, electricity, sound, energy of chemical reactions, magnetic attraction, energy of atomic reactions, and light. Definitions for a few of these types of energy are as follows:

Radiation - is the emission and propagation of energy in the form of ***electromagnetic waves***.

Chemical Energy - is energy used or released in chemical reactions.

Atomic Energy - is energy released from an atomic nucleus at the expense of its ***mass***.

Electromagnetic Energy - is energy stored in ***electromagnetic waves*** or ***radiation***. Energy is released when the waves are absorbed by a surface. Any object with a temperature above absolute zero (-273 degrees Celsius) emits this type of energy. The intensity of energy released is a function of the temperature of the radiating surface. The higher the temperature, the greater the quantity of energy released.

Electrical Energy - is energy resulting from the force between two objects having the physical property of charge.

Heat Energy - a form of energy representing aggregated internal energy of motions of ***atoms*** and ***molecules*** in a body.

On Earth, there are fundamentally three ways in which energy can be transferred from one place to another: ***conduction***, ***convection***, and ***radiation***. ***Conduction*** consists of energy transferred directly from molecule to molecule, and represents the flow of energy along a temperature gradient. ***Convection*** involves the transfer of energy by means of vertical mass motions of the medium through which heat is transferred (horizontal transfer is called ***advection***). In gases and liquids, this exchange by mass motions is commonly seen in rising bubbles known as ***convection currents***. Both conduction and convection depend on a material medium in order to operate. This medium can be gaseous, liquid, or solid. ***Radiation*** is the only means of energy transfer through space without the aid of a material medium, and is the major source of energy on the Earth.

Matter: Elements and Compounds

Matter is the material that makes up things in the Universe. All matter on the Earth is constructed of ***elements*** (see ***WebElements*** for the ***periodic table*** of elements). Chemists have described approximately 115 different elements. Each of these elements have distinct chemical characteristics. **Table 6a-1** lists some of the chemical characteristics for 48 common elements found in the Earth's ***continental crust***.

The smallest particle that exhibits the unique chemical characteristics of an element is known as an ***atom***. Atoms are composed of yet smaller particles known as ***protons***, ***neutrons***, and ***electrons***. A proton is a subatomic particle that has significant ***mass*** and contributes a single positive electrical charge to an atom. Neutrons also have significant mass but no electrical charge. Electrons are extremely light subatomic particles having a mass that is 1/1840 of a proton. Each electron also has a negative electrical charge.

Protons and neutrons make up the ***nucleus*** of an atom. As a result, most of an atom's ***mass*** is concentrated in the nucleus. Because protons are positively charged the nucleus has a positive charge equal to the number of these subatomic particles. Electrons are found orbiting outside the nucleus at various distances based on their energy level. The area occupied by the electrons has a negative charge equal to the number of these subatomic particles. If an atom has an equal number of electrons and protons its net electrical charge is zero. If there are more electrons than protons the charge of the atom is negative. Likewise, if there are less electrons than protons the charge of the atom is positive. In both cases, the exact charge is determined by subtracting protons from electrons. As a result, 4 protons minus 6 electrons give an atomic charge of -2.

The number of protons found in the nuclei of the different types of elements is unique and is referred to as the ***atomic number*** (**Table 6a-1**). All atoms of a specific element have the same number of protons in their nuclei. ***Atomic mass number*** is an atom's total number of neutrons and protons. Many elements have unequal numbers of neutrons and protons in their nucleus. An element's ***atomic weight*** refers to the total weight of neutrons, protons, and electrons. For example, the atomic weight of aluminum is 26.98 (**Table 6a-1**). Atomic number describes the number of protons found in an atom. For example, silver has an atomic number of 47 or 47 protons in its atom (**Table 6a-1**). Some elements can have variants containing different numbers of neutrons but similar numbers of protons. We call these variants ***isotopes***. Carbon has two isotopes. Its most common form is carbon-12 which has 6 protons plus 6 neutrons. About 99 % of the carbon on our planet is of this type. The isotope carbon-13 has 6 protons plus 7 neutrons. Carbon-14 is the rarest isotope of carbon containing 8 neutrons. Some isotopes are unstable and their nucleus tends to lose subatomic particles forming an element with a lower atomic mass. This process is known as ***radioactive decay***.

{PRIVATE**}Table 6a-1:** Characteristics of some of the common chemical elements found in the Earth's continental crust.

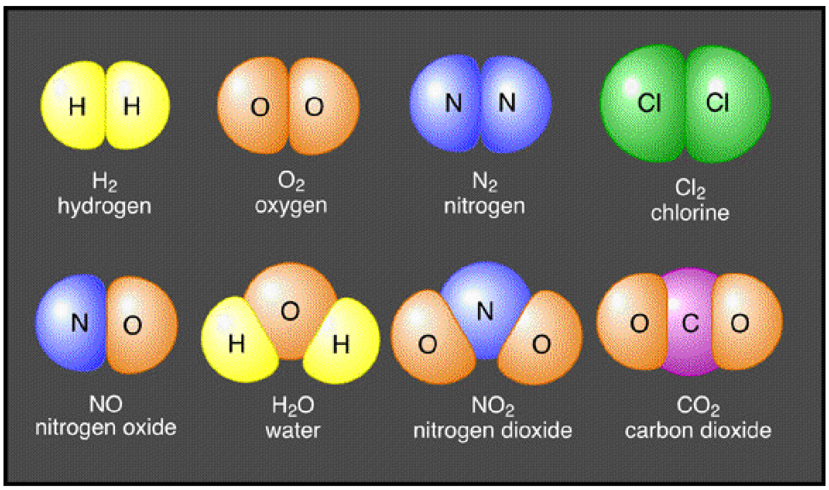
{PRIVATE**}Element**	Chemical Symbol	Atomic Number	Common Atomic	Atomic Weight	Percent in Continental	Required for all	Required for	Element Type	Moderately	Extremely

			Mass Number		Crust	Life	Some Lifeforms		Toxic	Toxic
Aluminum	Al	13	27	26.98	8.2300	-	**X**	**Metalloid**	-	-
Antimony	Sb	51	122	121.75	0.00002	-	-	**Metalloid**	-	-
Arsenic	As	33	75	74.92	0.00018	-	-	**Metalloid**	-	**X**
Barium	Ba	56	137	137.34	0.0425	-	-	**Metal**	-	-
Beryllium	Be	4	10	9.01	0.00028	-	-	**Metal**	-	**X**
Bismuth	Bi	83	209	208.98	0.000017	-	-	**Metal**	-	-
Boron	B	5	11	10.81	0.0010	-	-	**Metalloid**	-	-
Bromine	Br	35	80	79.91	0.00025	-	-	**Nonmetal**	-	-
Cadmium	Cd	48	112	112.40	0.00002	-	-	**Metal**	-	**X**
Calcium	Ca	20	40	40.08	4.1000	**X**	-	**Metal**	-	-
Carbon	C	6	12	12.01	0.0200	**X**	-	**Nonmetal**	-	-
Chlorine	Cl	17	35.5	35.45	0.0130	-	**X**	**Nonmetal**	**X**	-
Chromium	Cr	24	52	52.00	0.0100	-	-	**Metal**	**X**	-
Cobalt	Co	27	59	58.93	0.0025	-	**X**	**Metal**	-	-
Copper	Cu	29	63.5	63.54	0.0055	**X**	-	**Metal**	**X**	-
Fluorine	F	9	19	19.00	0.0625	-	**X**	**Nonmetal**	**X**	-
Gallium	Ga	31	70	69.72	0.0015	-	-	**Metal**	-	-
Germanium	Ge	32	73	72.59	0.00015	-	-	**Metalloid**	-	-
Gold	Au	79	197	196.97	0.0000004	-	-	**Metal**	-	-
Hydrogen	H	1	1	1.008	1.4000	**X**	-	**Nonmetal**	-	-
Iodine	I	53	127	126.90	0.00005	-	**X**	**Nonmetal**	-	-
Iron	Fe	26	56	55.85	5.6000	**X**	-	**Metal**	-	-
Lead	Pb	82	207	207.19	0.00125	-	-	**Metal**	-	**X**
Lithium	Li	3	6	6.94	0.0020	-	-	**Metal**	-	-
Magnesium	Mg	12	24	24.31	2.3000	**X**	-	**Metal**	-	-
Manganese	Mn	25	55	54.94	0.0950	**X**	-	**Metal**	-	-
Mercury	Hg	80	201	200.59	0.000008	-	-	**Metal**	-	**X**
Molybdenum	Mo	42	96	95.94	0.00015	**X**	-	**Metal**	-	-
Nickel	Ni	28	59	58.71	0.0075	-	-	**Metal**	-	**X**
Nitrogen	N	7	14	14.01	0.0020	**X**	-	**Nonmetal**	-	-
Oxygen	O	8	16	16.00	46.4000	**X**	-	**Nonmetal**	-	-
Palladium	Pd	46	106	106.40	0.000001	-	-	**Metal**	**X**	-
Phosphorus	P	15	31	30.97	0.1050	**X**	-	**Nonmetal**	-	-
Platinum	Pt	78	195	195.09	0.0000005	-	-	**Metal**	-	-
Potassium	K	19	39	39.10	2.1000	**X**	-	**Metal**	-	-
Rubidium	Rb	37	85.5	85.47	0.0090	-	-	**Metal**	-	-
Selenium	Se	34	79	78.96	0.000005	-	**X**	**Nonmetal**	**X**	-
Silicon	Si	14	28	28.09	28.2000	-	-	**Metalloid**	-	-
Silver	Ag	47	108	107.87	0.000007	-	-	**Metal**	-	**X**
Sodium	Na	11	23	22.99	2.4000	-	**X**	**Metal**	-	-
Sulfur	S	16	32	32.06	0.0260	**X**	-	**Nonmetal**	-	-
Thorium	Th	90	232	232.04	0.00096	-	-	-	-	-
Tin	Sn	50	119	118.69	0.00020	-	-	**Metal**	**X**	-
Titanium	Ti	22	48	47.90	0.5700	-	-	**Metal**	-	-
Tungsten	W	74	184	183.85	0.00015	-	-	**Metal**	-	-
Uranium	U	92	238	238.03	0.00027	-	-	-	-	-
Vanadium	V	23	51	50.94	0.0135	-	**X**	**Metal**	**X**	-
Zinc	Zn	30	65	65.37	0.0070	**X**	-	**Metal**	-	-

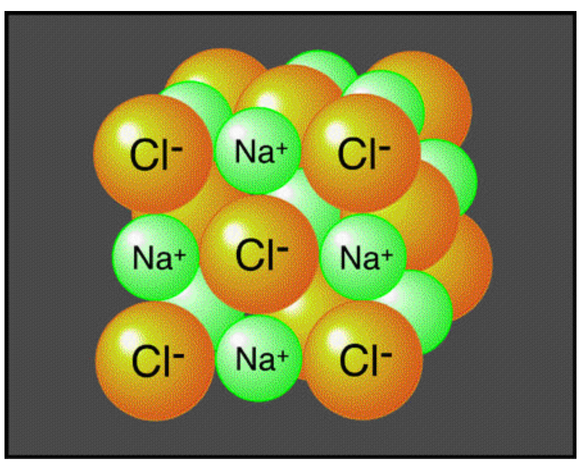
Elements can be classified as being either **metals**, **nonmetals**, or **metalloids (Table 6a-1)**. **Metals** are elements that usually conduct heat and electricity and are shiny. **Nonmetals** do not conduct electricity that well and are normally not shiny. **Metalloids** have characteristics that are in between metals and nonmetals.

Elements with a net positive or negative charge are called ***ions***. Chemists indicate the number of positive or negative charges on an ion using a superscript after the element's symbol. For example, calcium has two positive charges and is written as Ca^{2+}. Some common negatively charged ions include nitrate (NO_3^-), sulfate (SO_4^{2-}), and phosphate (PO_4^{3-}).

Positive and negative ions are electrically attracted to each other. This mutual attraction allows for the bonding of atoms to occur forming structures of matter that are larger than just one atom. When similar atoms bond together they construct ***molecules***. Atoms of different elements joined together form ***compounds*** (**Figure 6a-1**). Sodium chloride (or table salt), is an ionic compound consisting of sodium (Na^+) and chloride (Cl^-). In nature, it forms as a three-dimensional array of oppositely charged ions (**Figure 6a-2**). Many of the Earth's substances have a molecular structure similar to sodium chloride.



{PRIVATE}Figure 6a-1: Some common molecules and compounds. The molecules in the top row bond with each other by sharing electrons. The compounds in the bottom row also share electrons. However, these joins are called ionic bonds.



{PRIVATE}Figure 6a-2: Atomic representation of sodium chloride or table salt. This compound forms in nature as a highly ordered, three-dimensional network of oppositely charged ions. The bonds that form between the sodium (Na^+) and (Cl^-) chloride ions give this compound great internal strength allowing it to form large crystals.

Inorganic vs. Organic

Compounds and molecules constructed in living tissues are commonly called ***organic***. Forms of matter not formed by living things are termed ***inorganic***. Organisms like ***autotrophs*** usually create organic matter by consuming inorganic molecules and compounds from the lithosphere, hydrosphere, and atmosphere. An example of an autotroph is any photosynthesizing plant. ***Heterotrophs*** consume and ***assimilate*** other living things to create their organic matter. ***Herbivores*** and ***carnivores*** are examples of heterotrophs.

There are four general categories of organic compounds: ***lipids***, ***carbohydrates***, ***proteins***, and ***nucleic acids***.

Lipids - are composed of carbon atoms that have two hydrogen atoms attached. Lipids are commonly known as fats and oils, and belong to the family of molecules known as hydrocarbons.

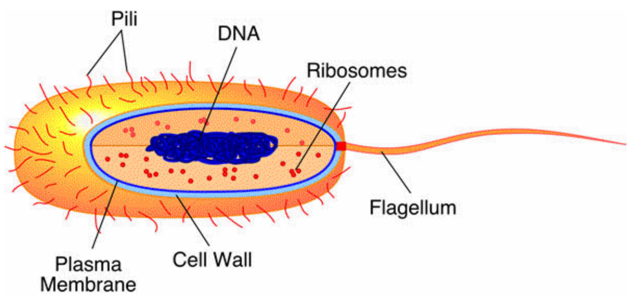
Carbohydrates - are composed of carbon, oxygen, and hydrogen atoms. Some examples are sugars, starch, and cellulose.

Proteins - are organic compounds that are made primarily of carbon, hydrogen, nitrogen, and some other minor elements that are arranged into 20 different compounds known as ***amino acids***.

Nucleic Acids - are composed primarily of different combinations of carbon, hydrogen, nitrogen, oxygen, and phosphorus. They are very complex compounds being created by the atomic linking of thousands of individual atoms. ***DNA*** or ***deoxyribonucleic acid***, the genetic blueprint of life, is an example of a nucleic acid.

Cells

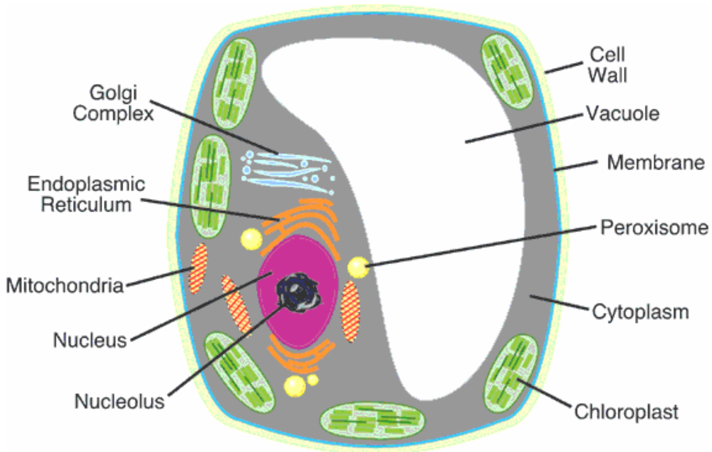
All organisms are composed of one or more of ***cells***. Cells are the smallest self-functioning unit found in living organisms. Cells are also where the processes of ***metabolism*** and ***heredity*** occur in an organism. Cells arise by the cellular division of a previously existing cell. Biologists have differentiated two basic types of cells in organisms. ***Bacteria***, ***archaea***, and ***cyanobacteria*** have cells that are quite uncomplicated in terms of structure and function. Quite simply, they lack internal organization. These cells are commonly known as ***prokaryotes*** (**Figure 6a-3**).



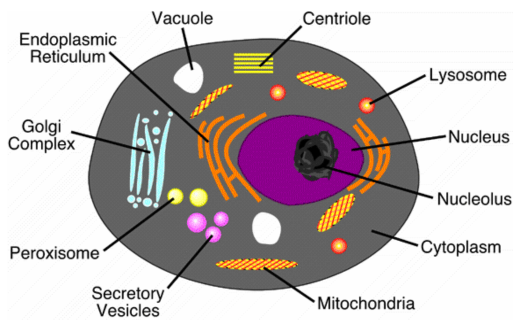
{PRIVATE**}Figure 6a-3:** Typical ***prokaryote*** cell. These cells are about 1 to 10 micrometers in size. The cell is encased by rigid cell wall and a plasma membrane. Within the cell, the two most obvious structures are ribosomes and ***DNA***. The DNA is not bounded by a membrane. Many prokaryote cells also have a flagellum which is used for movement.

The cells of plants and animals are more complex than those of bacteria, archaea, and cyanobacteria. We call these type of cells ***eukaryotes***. Eukaryotic cells have a membrane enclosed nucleous which contains the organism's ***DNA***. Plant and animal cells also contain a variety of membrane-bound structures known as ***organelles***. **Figures 6a-4** and **6a-5** describe the various structures found in typical plant and animals cells.

Eukaryotic cells also show great variation in terms of size. White blood cells of mammals are some of the smallest eukaryotic cells with a diameter between 3 to 4 micrometers. Ostrich ovum are very large cells with a diameter of about 100 micrometers.



{PRIVATE}**Figure 6a-4:** Typical plant cell. Plants cells differ from animal cells in the following ways: they have a cell wall, chloroplasts, and they often contain a large central vacuole.



{PRIVATE}**Figure 6a-5:** Typical animal cell.

Internally, cells contain specialized structures known as ***organelles*** that carry out distinct cellular functions. Within these structures ***enzymes***, a type of protein, are used to facilitate and regulate various chemical reactions. **Table 6a-2** describes the function of a variety of cell structures including many organelles.

{PRIVATE}**Table 6a-2:** Description and function of common cell structures.

{PRIVATE}**Structure**	**Description**	**Function**
Cell Wall	Outer layer on a cell composed of cellulose or other complex carbohydrates.	Helps to support and protect the cell.
Plasma Membrane	A layer composed of lipids and proteins that controls the permeability of the cell to water and dissolved substances.	Regulates the movement of material into and out of the cell.
Flagellum (Flagella pl.)	Threadlike organelle that extends from the surface of the cell. Found in both prokaryotes and eukaryotes.	Used for movement of the cell or to move fluids over the cell's surface for absorption.
Pilus (Pili pl.)	Hollow, hairlike structures made of protein found on prokaryote cells.	Allows cell to attach itself to another cell.
Ribosomes	Tiny, complex structures composed of protein and RNA. Often attached to endoplasmic reticulum.	Ribosomes are involved in protein synthesis.
Endoplasmic Reticulum	Extensive system of internal membranes.	Forms compartments to isolate cell substances.
Nucleus	Double membrane structure that encases chromosomes.	Control center of the cell which directs protein synthesis and cell reproduction.
Chromosomes	Long strands of DNA.	Store hereditary information.
Nucleolus	Aggregations of rRNA and ribosomal proteins.	Area were ribosomes are manufactured.
Golgi Complex	Flattened stacks of membranes.	Used in the collection, packaging, and distribution of synthesized molecules.

Peroxisomes	Membrane confined spherical body about 0.2 to 0.5 micrometers in diameter.	Formed by the endoplasmic reticulum. Converts fats into carbohydrates. Detoxifies potentially harmful oxidants.
Lysosomes	Membrane confined spherical body about 0.2 to 0.5 micrometers in diameter.	Formed by the golgi complex. Contains digestive enzymes for braking down old cellular components.
Centrioles	Long hollow tubes composed of protein. Not found in plant cells.	Influence cell shape, move chromosomes during reproductive division, and are the internal structure for flagellum.
Secretory Vesicles	Membrane enclosed sack created at the golgi complex.	These structures contain cell secretions, like hormones and neurotransmitters. The secretory vesicles are then transported to the cell surface where they are release to the environment outside the cell.
Vacuole	Voids within the cytoplasm. Quite large in plant cells.	Used to store water and waste products.
Cytoplasm	Semifluid mixture that occupies most of the cell's interior. Contains sugars, amino acids, and proteins. Also, contains a protein fiber network.	Medium in which organelles and other internal structures exist in. Fiber network support the shape of the cell and anchor organelles to fixed positions.
Mitochondria	Elongated structures about 1 to 3 micrometers long. Resemble aerobic bacteria.	Structure which converts sugar into energy through oxidation.
Chloroplasts	Elongated structures with vesicles containing chlorophyll.	Site of photosynthesis.

Cells can also be classified according to how they obtain their energy. Some cells have the ability to use light or chemical energy found in the outside environment to manufacture their own sugars, fats, and proteins. We call these types of cells ***autotrophs***. All 400,000 species of plants and a few species of bacteria use sunlight and the process of ***photosynthesis*** to obtain their energy. Some bacteria breakdown molecules found in the environment to release chemical energy to sustain their life. Organisms can also obtain their energy by consuming other organisms. These organisms are called ***heterotrophs***. Heterotrophs include most types of bacteria and all of the animal and fungi species.

Some organisms consist of just one cell (***bacteria***, ***algae***, and ***protozoa***). However, most organisms are **multi-cellular**. Within multi-cellular organisms, groups of cells can become specialized to carry out a certain function. We call these functional groups of cells a ***tissue***. An ***organ*** is a structure composed of several different types of tissues. Organs also have a specific structure and a particular function.

(b) Measurement of Energy

{PRIVATE**}**In the previous section, we developed the concept of energy. We now must be able to measure and quantify it, using a **standard** set of **units**. Worldwide, two systems of units of measurement are commoly used today: the **Metric System** (Systeme International) and the **British System**.

The units of energy described in these systems are derived from a technical definition of energy used by physicists. This definition suggests that energy can be represented by the following simple equation:

Work = Force x Distance

Similar to the definition given in the previous topic, physicists view energy as the ability to do work. However, they define **work** as a **force** applied to some form of ***matter*** (object) multiplied by the **distance** that this object travels. Physicists commonly describe force with a unit of measurement known as a ***newton*** (after Sir Isaac Newton). A **newton** is equal to the force needed to accelerate (move) a mass weighting one kilogram one meter in one second in a ***vacuum*** with no ***friction***. The **work** or energy required to move an object with the **force** of one **newton** over a **distance** of one meter is called a ***joule***.

Some other definitions for the energy measurement units that you may come across in this textbook are as follows:

Calorie - equals the amount of ***heat*** required to raise 1 gram of pure water from 14.5 to 15.5 degrees Celsius at ***standard atmospheric pressure***. 1 calorie is equal to 4.1855 joules. The abreviation for calorie is **cal**. A kilocalorie, abbreviated **kcal**, is equal to a 1000 calories. 1 kilocalorie is equal to 4185 joules.

Btu - also called ***British thermal unit*** is the amount of energy required to raise the temperature of one pound of water one degree ***Fahrenheit***.

Langley (**ly min**$^{-1}$) - unit of the intensity of ***radiation*** measured per minute and equal to one calorie.

Watt (**Wm**$^{-2}$) - a metric unit of measurement of the intensity of ***radiation*** in watts over a square meter surface. One watt is equal to one ***joule*** of work per second. A kilowatt (**kW**) is the same as 1000 watts. One langley min^{-1} is equivalent to 697.3 Wm^{-2}.

(c) Energy, Temperature, and Heat

{PRIVATE}So far we have learned that energy can take on many forms. One important form of energy, relative to life on Earth, is ***kinetic energy***. Simply defined, kinetic energy is the energy of **motion**. The amount of kinetic energy that a body possesses is dependent on the speed of its motion and its ***mass***. At the **atomic scale**, the kinetic energy of atoms and molecules is sometimes referred to as ***heat energy***.

Kinetic energy is also related to the concept of ***temperature***. Temperature is defined as the measure of the average speed of atoms and molecules. The higher the temperature, the faster these particles of ***matter*** move. At a temperature of - 273.15 degrees Celsius (***absolute zero***) all atomic motion stops. ***Heat*** is often defined as energy in the process of being transferred from one object to another because of the temperature difference between them. In the atmosphere, heat is commonly transferred by ***conduction***, ***convection***, ***advection***, and ***radiation***.

Some other important definitions related to energy, temperature, and heat are:

Heat Capacity - of a substance is the ratio of the amount of ***heat energy*** absorbed by that substance to its corresponding temperature rise.

Specific Heat - is the heat capacity of a unit mass of a substance or heat needed to raise the temperature of 1 gram (g) of a substance 1 degree Celsius.

Sensible Heat - is heat that can be measured by a ***thermometer***, and thus sensed by humans. Several different scales of measurement exist for measuring sensible heat. The most common are: ***Celsius scale***, ***Fahrenheit scale***, and the ***Kelvin scale***.

Latent Heat - is the energy required to change a substance to a higher state of matter. This same energy is released from the substance when the change of state is reversed. The **diagram** below describes the various exchanges of heat involved with 1 gram of water.

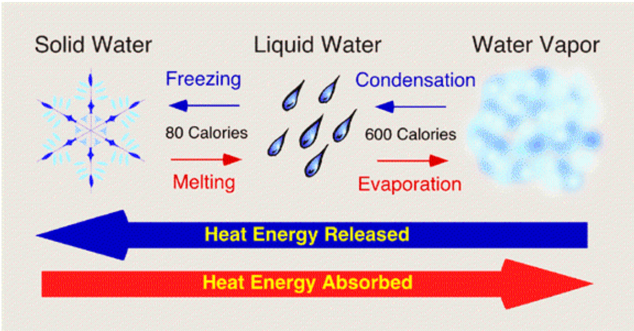
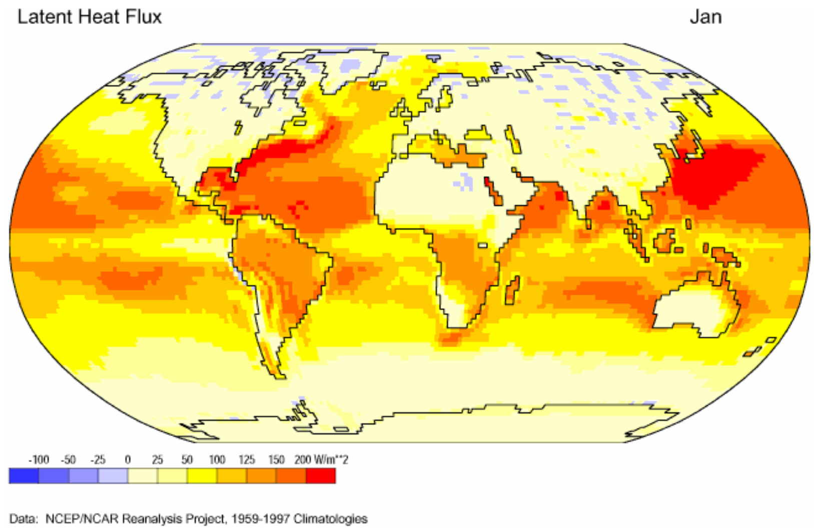
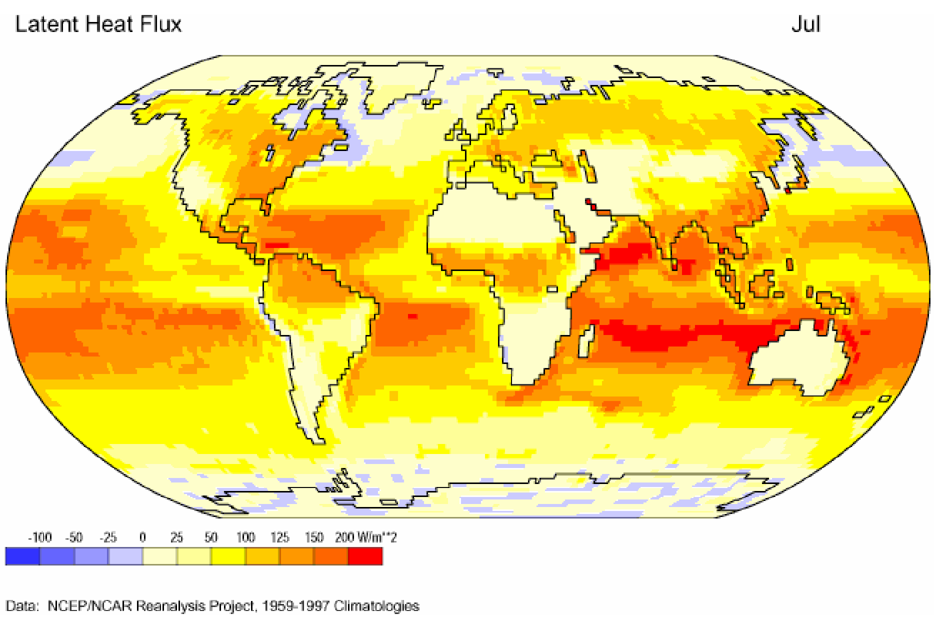


Figure 6c-1: Latent heat exchanges of energy involved with the phase changes of water.

Figures 6c-2 and **6c-3** show the net absorption and release of latent heat energy for the Earth's surface for January and July, respectively. The highest values of flux or flow occur near the subtropical oceans where high temperatures and a plentiful supply of water encourage the evaporation of water. Negative values of latent heat flux indicate a net release of latent energy back into the environment because of the condensation or freezing of water. Values of latent heat flux are generally low over landmasses because of a limited supply of water at the ground surface.



{PRIVATE**}Figure 6c-2:** Mean January latent heat flux for the Earth's surface, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).



{PRIVATE**}Figure 6c-3:** Mean July latent heat flux for the Earth's surface, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

(d) Energy and Life

{PRIVATE**}**The capture and use of energy in living systems is dominated by two processes: ***photosynthesis*** and ***respiration***. Through these two processes living organisms are able to capture and use all of the energy they require for their activities.

Photosynthesis

Plants can capture the electromagnetic energy from the sun by a chemical process called ***photosynthesis***. This chemical reaction can be described by the following simple equation:

$6CO_2 + 6H_2O$ + light energy = $C_6H_{12}O_6 + 6O_2$

The product of photosynthesis is the ***carbohydrate glucose*** and oxygen which is released into the atmosphere. All of the sugar ***glucose*** is produced in the specialized photosynthetic cells of plants and some other organisms. Glucose is produced by chemically combining carbon dioxide and water with sunlight. This chemical reaction is catalyzed by ***chlorophyll*** acting in concert with other pigment, lipid, sugars, protein, and nucleic acid molecules. Sugars created in photosynthesis can be later converted by the plant to starch for storage, or it can be combined with other sugar molecules to form specialized carbohydrates such as ***cellulose***, or it can be combined with other nutrients such as nitrogen, phosphorus, and sulfur, to build complex molecules such as ***proteins*** and ***nucleic acids***.

Because all the energy fixed by a plant is converted into sugar, it is theoretically possible to determine a plant's energy uptake by measuring the amount of sugar produced. This quantity is called ***gross primary productivity***. Measurements of the buildup of sugar in the plant reflect ***gross primary productivity*** less ***respiration***, or ***net primary productivity***.

In general, animals cannot produce their own energy via photosynthesis. Instead, they capture their energy by the consumption and ***assimilation*** of the biomass of plants or other animals. Thus, animals get the energy they need for maintenance of their bodies tissues, growth, and reproduction indirectly from photosynthetic organisms.

Respiration

The ***oxidation*** of sugar by organisms is called ***respiration***. This process occurs in both plants and animals. In most organisms, respiration releases the energy required for all metabolic processes. This chemical reaction can be described by the following simple equation:

$C_6H_{12}O_6 + 6O_2 = 6CO_2 + 6H_2O$ + released energy

One of the products of respiration is energy, which is released via the chemical decomposition of ***glucose***. Other products of this chemical reaction are carbon dioxide (**CO_2**) and water (**H_2O**).

(e) Laws of Thermodynamics

{PRIVATE**}**The field of thermodynamics studies the behavior of ***energy*** flow in natural systems. From this study, a number of physical laws have been established. The **laws of thermodynamics** describe some of the fundamental truths of thermodynamics observed in our ***Universe***. Understanding these laws is important to students of Physical Geography because many of the processes studied involve the flow of energy.

First Law of Thermodynamics

The first law of thermodynamics is often called the **Law of Conservation of Energy**. This law suggests that ***energy*** can be transferred from one ***system*** to another in many forms. However, it can not be **created** nor **destroyed**. Thus, the total amount of energy available in the Universe is constant. Einstein's famous equation (written **below**) describes the relationship between energy and ***matter***:

$E = MC^2$

In the equation above, ***energy*** (**E**) is equal to ***matter*** (**M**) times the square of a constant (**C**). Einstein suggested that energy and matter are interchangeable. His equation also suggests that the quantity of energy and matter in the Universe is fixed.

Second Law of Thermodynamics

Heat can never pass spontaneously from a colder to a hotter body. As a result of this fact, natural processes that involve energy transfer must have one direction, and all natural processes are irreversible. This law also predicts that the ***entropy*** of an isolated system always increases with time. Entropy is the measure of the disorder or randomness of ***energy*** and ***matter*** in a system. Because of the second law of thermodynamics both energy and matter in the ***Universe*** are becoming less useful as time goes on. Perfect order in the Universe occurred the instance after the ***Big Bang*** when energy and matter and all of the forces of the Universe were unified.

Third Law of Thermodynamics

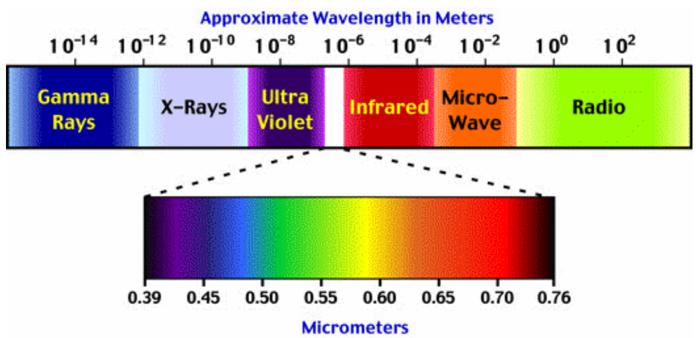
The third law of thermodynamics states that if all the **thermal motion** of ***molecules*** (***kinetic energy***) could be removed, a state called ***absolute zero*** would occur. Absolute zero results in a ***temperature*** of 0 ***Kelvin*** or -273.15 degrees ***Celsius***.

Absolute Zero = 0 Kelvin = -273.15 degrees Celsius

The Universe will attain absolute zero when all energy and matter is randomly dis tributed across space. The current temperature of empty space in the Universe is about 2.7 Kelvins.

(f) The Nature of Radiation

{PRIVATE**}**All objects above the ***temperature*** of ***absolute zero*** (-273.15 degrees Celsius) ***radiate energy*** to their surrounding environment. This energy, or radiation, is emitted as ***electromagnetic waves*** that travel at the ***speed of light***. Many different types of radiation have been identified. Each of these types is defined by its ***wavelength***. The wavelength of electromagnetic radiation can vary from being infinitely short to infinitely long (**Figure 6f-1**).



{PRIVATE**}Figure 6f-1:** Some of the various types of electromagnetic radiation as defined by wavelength. Visible light has a spectrum that ranges from 0.40 to 0.71 micrometers (µm).

Visible ***light*** is a form of electromagnetic radiation that can be perceived by our eyes. Light has a ***wavelength*** of between 0.40 to 0.71 micrometers (µm). **Figure 6f-1** illustrates that various spectral color bands that make up light. The sun emits only a portion (44 %) of its radiation in zone. Solar radiation spans a spectrum from approximately 0.1 to 4.0 micrometers. The band from 0.1 to 0.4 micrometers is called ***ultraviolet radiation***. About 7 % of the sun's emission is in this wavelength band. About 48 % of the sun's radiation falls in the region between 0.71 to 4.0 micrometers. This band is called the near (0.71 to 1.5 micrometers) and far ***infrared*** (1.5 to 4.0 micrometers).

The amount of electromagnetic radiation emitted by a body is directly related to its temperature. If the body is a perfect emitter (***black body***), the amount of radiation given off is proportional to the 4th power of its temperature as measured in ***Kelvin*** units. This natural phenomenon is described by the ***Stephan-Boltzmann Law***. The following simple equation describes this law mathematically:

$$E^* = \sigma T^4$$

where σ (sigma) = $5.67 \times 10^{-8}\ Wm^{-2}\ K^{-4}$

and T is the temperature in Kelvin

According to the Stephan-Boltzmann equation, a small increase in the temperature of a radiating body results in a large amount of additional radiation being emitted.

In general, good emitters of radiation are also good absorbers of radiation at specific ***wavelength*** bands. This is especially true of gases and is responsible for the Earth's ***greenhouse effect***. Likewise, weak emitters of radiation are also weak absorbers of radiation at specific wavelength bands. This fact is referred to as ***Kirchhoff's Law***.

Some objects in nature have almost completely perfect abilities to absorb and emit radiation. We call these objects ***black bodies***. The radiation characteristics of the sun and the Earth are very close to being black bodies.

The ***wavelength*** of **maximum emission** of any body is inversely proportional to its absolute temperature. Thus, the higher the temperature, the shorter the wavelength of maximum emission. This phenomenon is often called ***Wien's Law***. The following equation describes this law:

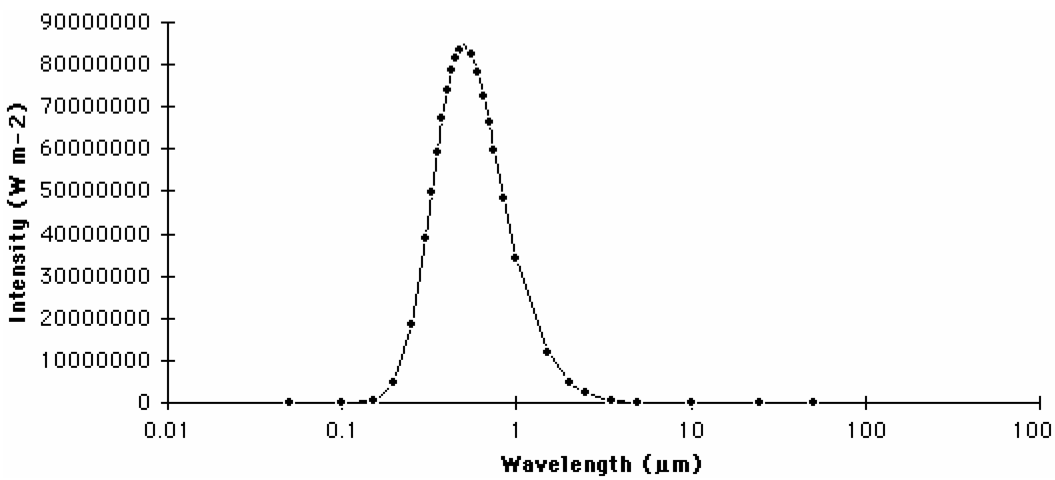
$$\lambda_{max} = C/T$$

where C is a constant equal to 2897
and T is the temperature in Kelvin

Wien's law suggests that as the temperature of a body increases, the wavelength of maximum emission becomes smaller. According to the above equation the wavelength of maximum emission for the sun (5800 Kelvins) is about 0.5 micrometers, while the wavelength of maximum emission for the Earth (288 Kelvins) is approximately 10.0 micrometers.

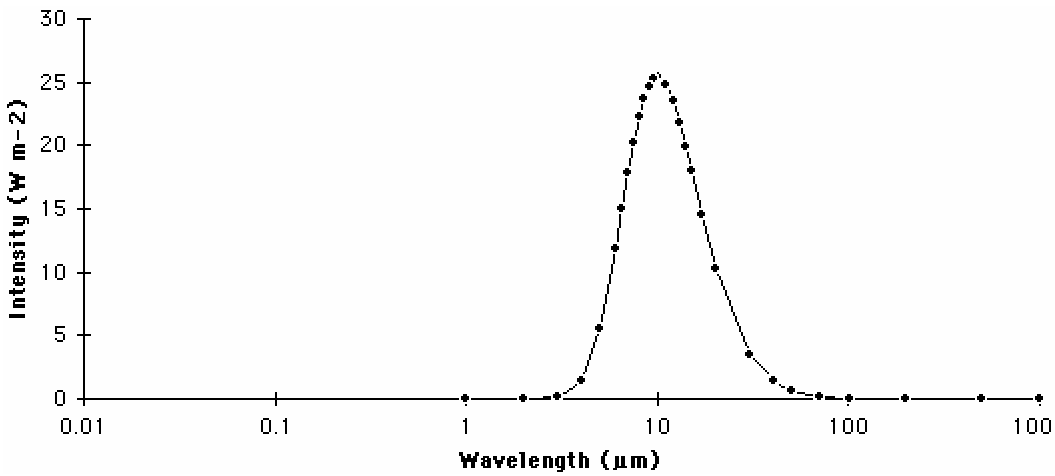
A graph that describes the quantity of radiation that is emitted from a body at particular wavelengths is commonly called a ***spectrum***. The following two **graphs** describe the spectrums for the sun and Earth.

SUN



{PRIVATE**}Figure 6f-2:** Spectrum of the sun. The sun emits most of its radiation in a wavelength band between 0.1 and 4.0 micrometers (µm).

EARTH



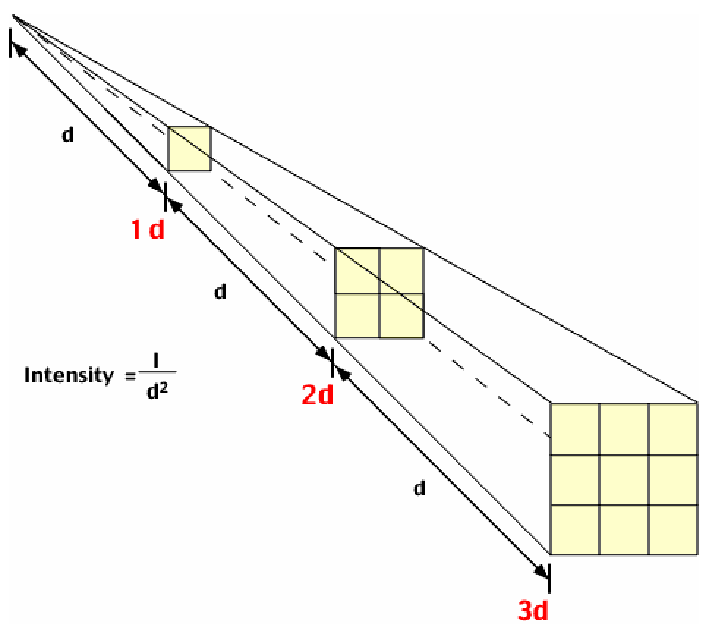
{PRIVATE}Figure 6f-3: Spectrum of the Earth. The Earth emits most of its radiation in a wavelength band between 0.5 and 30.0 micrometers (µm).

The above graphs illustrate two important points concerning the relationship between the temperature of a body and its emissions of electromagnetic radiation:

The amount of radiation emitted from a body increases exponentially with a linear rise in temperature (see above **Stephan-Boltzmann's Law**).

The average wavelength of electromagnetic emissions becomes shorter with increasing temperature (see above **Wien's Law**).

Finally, the amount of radiation passing through a specific area is inversely proportional to the square of the distance of that area from the energy source. This phenomenon is called the ***Inverse Square Law***. Using this law we can model the effect that distance traveled has on the intensity of emitted radiation from a body like the sun. **Figure 6f-4** suggests that the intensity of radiation emitted by a body quickly diminishes with distance in a nonlinear fashion.



{PRIVATE}Figure 6f-4: Diagram illustrating the diffusion of radiation due to the Inverse Square Law.

Mathematically, the **Inverse Square Law** is described by the equation: Intensity = I/d^2

where **I** is the intensity of the radiation at 1d (see above **diagram**) and **d** is the distance traveled.

Let us try this equation out. For example, what would be the intensity of emitted radiation traveling **two** units of distance if the intensity at 1d = 90,000 units?

Intensity = 90,000/2

Intensity = 90,000/4

Intensity = 22,500 units

What would be the intensity of emitted radiation traveling **three** units of distance if the intensity at 1d = 90,000 units?

Intensity = 90,000/3

Intensity = 90,000/9

Intensity = 10,000 units

What would be the intensity of emitted radiation traveling **four** units of distance if the intensity at 1d = 90,000 units?

Intensity = 90,000/4

Intensity = 90,000/16

Intensity = 5625 units

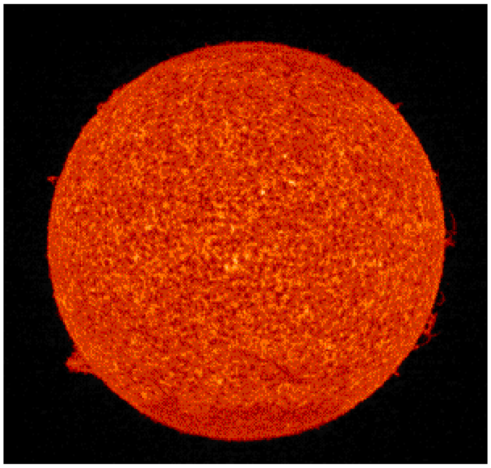
Note that the decrease in intensity with distance is not linear when graphed (**Figure 6f-5**)!

Figure 6f-5: Reduction in intensity of radiation with distance traveled.

(g) The Solar Source of the Earth's Energy

{PRIVATE}Almost all of the ***energy*** that drives the various ***systems*** (climate systems, ecosystems, hydrologic systems, etc.) found on the Earth originates from the ***sun***. Solar energy is created at the core of the sun when hydrogen atoms are fused into helium by ***nuclear fusion***. For each second of this nuclear process, 700 million tons of hydrogen are converted into 695 million tons of helium. The remaining 5 million tons are turned into ***electromagnetic energy*** that radiates from the sun's surface out into space.

The radiative surface of the sun, or ***photosphere***, has an average temperature of about 5800 Kelvins. Most of the electromagnetic radiation emitted from the sun's surface lies in the visible band centered at 0.5 µm. The total quantity of energy emitted from the sun's surface is approximately 63,000,000 ***Watts*** per square meter (Wm^{-2}).



{PRIVATE}**Figure 6g-1:** The sun observed by SUMER instrument on the SOHO satellite on March 2-4, 1996. (**Source:** *SOHO* - SUMER Instrument).

The energy emitted by the sun passes through space until it is intercepted by planets and other celestial objects. The intensity of solar radiation striking these objects is determined by a physical law known as the ***Inverse Square Law*** (see **topic *6f***). This law merely states that the intensity of the radiation emitted from the sun varies with the squared distance from the source. As a result of this law, if the intensity of radiation at a given distance is one unit, at twice the distance the intensity will become only one-quarter. At three times the distance, the intensity will become only one-ninth of its original intensity at a distance of one unit, and so on.

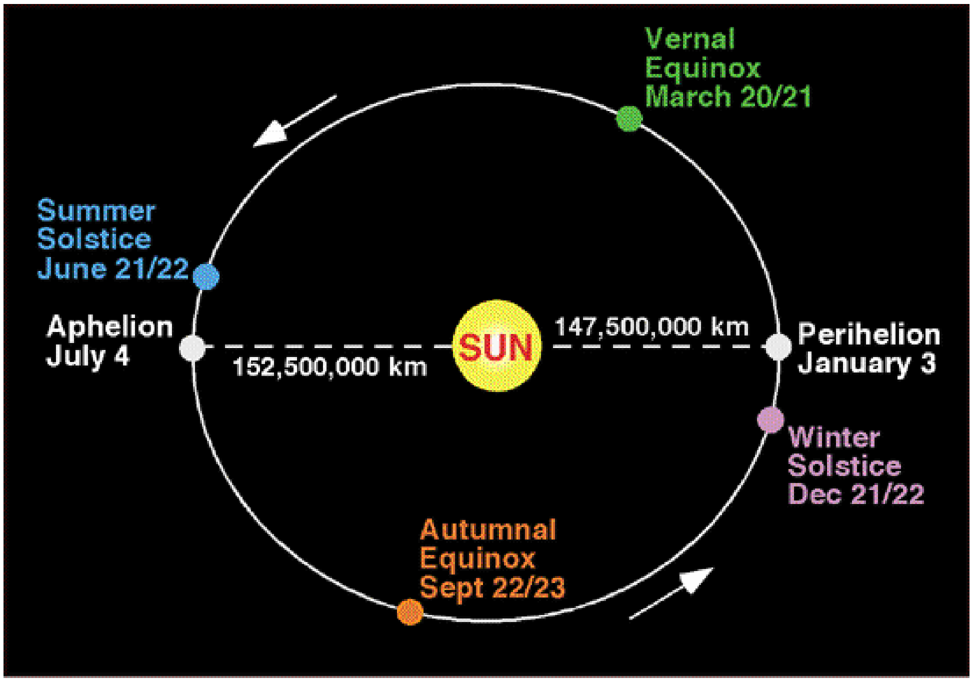
Given the amount of energy radiated by the sun and the average Earth-sun distance of 149.5 million kilometers, the amount of radiation intercepted by the outer limits of the atmosphere can be calculated to be around 1370 Wm^{-2}. For general purposes, the energy output of the sun can be considered constant. This of course is not entirely true. Scientists have shown that the output of the sun is temporally variable. Some researchers have also suggested that the increase in the ***average global temperature*** over the last century may have been solar in origin. This statement, however, is difficult to prove because accurate data on solar output of radiation only goes back to about 1978 (see ***link***).

(h) Earth-Sun Geometry

{PRIVATE}Earth Rotation and Revolution

The term ***Earth rotation*** refers to the spinning of the Earth on its axis. One rotation takes exactly twenty-four hours and is called a ***mean solar day***. If you could look down at the Earth's North Pole from space you would notice that the direction of rotation is counterclockwise. The opposite is true if you viewed the Earth from the South Pole.

The orbit of the Earth around the sun is called ***Earth revolution***. This celestial motion takes 365 1/4 days to complete one cycle. Further, the Earth's orbit around the sun is not circular, but **elliptical** (see **Figure 6h-1**). An elliptical orbit causes the Earth's distance from the sun to vary annually. However, this phenomenon does **not** cause the seasons! This annual variation in the distance from the sun does influence the amount of solar radiation intercepted by the Earth by approximately 6 %. On January 3rd, ***perihelion***, the Earth is closest to the sun (147.5 million kilometers). The Earth is farthest from the sun on July 4th, or ***aphelion***. The average distance of the Earth from the sun over a one year period is 150 million kilometers.

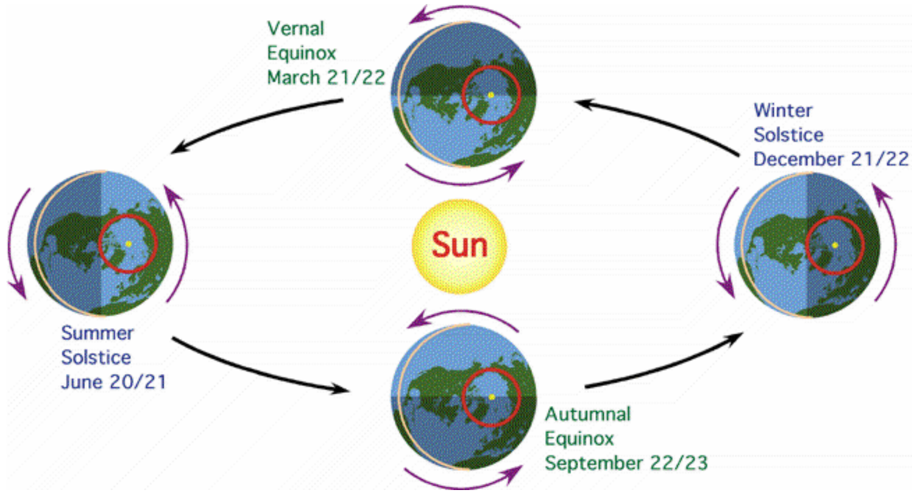


{PRIVATE}**Figure 6h-1:** Position of the equinoxes, solstices, aphelion, and perihelion on the Earth's orbit.

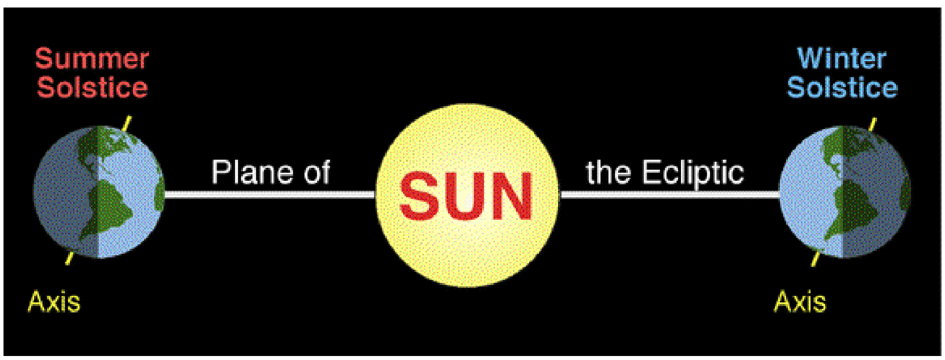
Tilt of the Earth's Axis

The Earth's axis is not **perpendicular** to the ***plane of the ecliptic***, but inclined at a fixed angle of 23.5 degrees. Moreover, the northern end of the Earth's axis always points to the same place in space (**North Star**). The following link shows an animation of the Earth revolving around the sun: ***Earth Revolution Animation - Figure 6h-5***. In this animation the Earth's axis is colored red. Note that the angle of the Earth's axis in relation to the plane of the ecliptic remains unchanged. However, the relative position of the Earth's axis to the sun does change during this cycle (**Figure 6h-2**). This circumstance is responsible for the annual changes in the height of the sun above the ***horizon***. It also causes the ***seasons***, by controlling the **intensity** and **duration** of sunlight received by locations on the Earth.

On June 21 or 22, the ***summer solstice***, the Earth is positioned in its orbit so that the North Pole is leaning 23.5 degrees toward the sun (**Figures 6h-2**, **6h-3** and see ***animation - Figure 6h-6***). During the summer solstice, all locations North of the equator have day lengths greater than twelve hours, while all locations South of the equator have day lengths less than twelve hours (see ***Table 6h-2***). On December 21 or 22, the ***winter solstice***, the Earth is positioned so that the South Pole is leaning 23.5 degrees toward the sun (**Figures 6h-2**, **6h-3** and see ***animation - Figure 6h-7***). During the winter solstice, all locations North of the equator have day lengths less than twelve hours, while all locations South of the equator have day lengths greater than twelve hours (see ***Table 6h-2***).

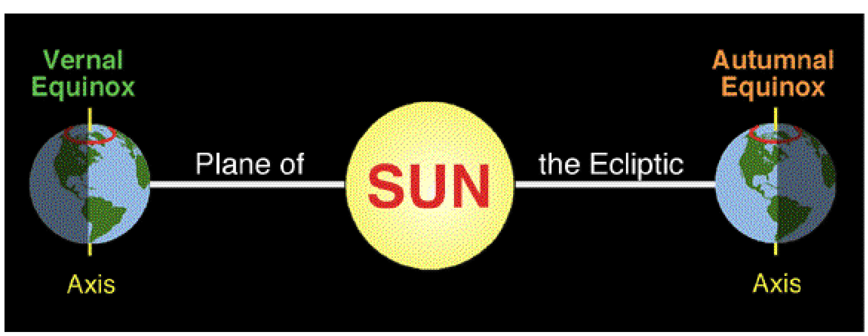


{PRIVATE}**Figure 6h-2:** Annual change in the position of the Earth in its revolution around the sun. In this graphic, we are viewing the Earth from a position in space that is above the North Pole (yellow dot) at the summer solstice, the winter solstice, and the two equinoxes. Note how the position of the North Pole on the Earth's surface does not change. However, its position relative to the sun does change and this shift is responsible for the seasons. The red circle on each of the Earths represents the Arctic Circle (66.5 degrees N). During the summer solstice, the area above the Arctic Circle is experiencing 24 hours of daylight because the North Pole is tilted 23.5 degrees toward the sun. The Arctic Circle experiences 24 hours of night when the North Pole is tilted 23.5 degrees away from the sun in the winter solstice. During the two equinoxes, the circle of illumination cuts through the polar axis and all locations on the Earth experience 12 hours of day and night.



{PRIVATE}**Figure 6h-3:** During the summer solstice the Earth's North Pole is tilted 23.5 degrees towards the sun relative to the circle of illumination. This phenomenon keeps all places above a latitude of 66.5 degrees N in 24 hours of sunlight, while locations below a latitude of 66.5 degrees S are in darkness. The North Pole is tilted 23.5 degrees away from the sun relative to the circle of illumination during the winter solstice. On this date, all places above a latitude of 66.5 degrees N are now in darkness, while locations below a latitude of 66.5 degrees S receive 24 hours of daylight.

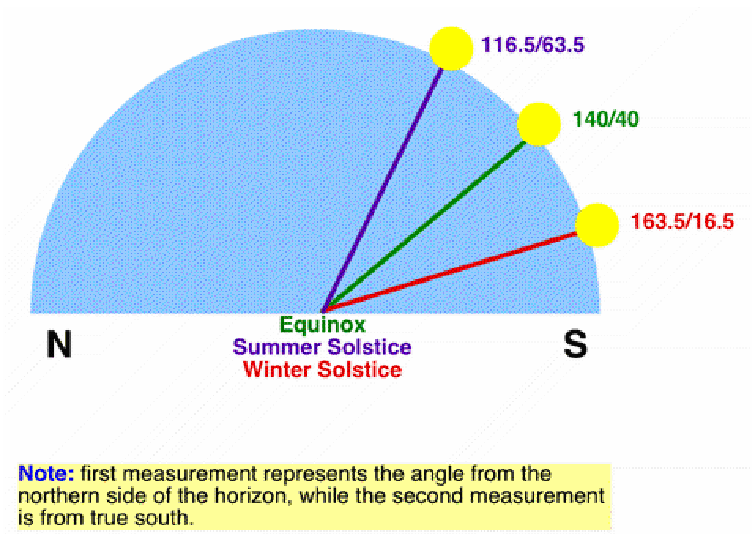
On September 22 or 23, the ***autumnal equinox***, neither pole is tilted toward the sun (**Figures 6h-2**, **6h-4** and see ***animation - Figure 6h-8***). March 20 or 21 marks the arrival of the ***spring*** or ***vernal equinox*** when once again the poles are not tilted toward the sun. Day lengths on both of these days, regardless of latitude, are exactly 12 hours.



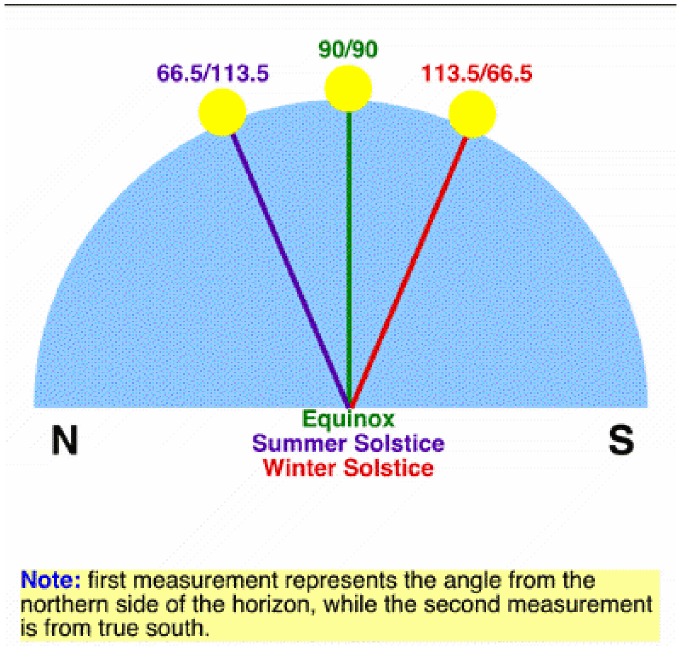
{PRIVATE}**Figure 6h-4:** During the equinoxes, the axis of the Earth is not tilted toward or away from the sun and the circle of illumination cuts through the poles. This situation does not suggest that the 23.5 degree tilt of the Earth no longer exists. The vantage point of this graphic shows that the Earth's axis is inclined 23.5 degrees toward the viewer for both dates (see **Figure 6h-2**). The red circles shown in the graphic are the Arctic Circle.

Axis Tilt and Solar Altitude

The annual change in the relative position of the Earth's axis in relationship to the sun causes the height of the sun (***solar altitude***) to vary in our skies. The total variation in maximum solar altitude for any location on the Earth over a one year period is 47 degrees (2 x 23.5 = 47). For example, at 50 degrees North maximum solar altitude varies from 63.5 degrees on the summer solstice to 16.5 degrees on the winter solstice (**Figure 6h-9**). Maximum solar height at the equator goes from 66.5 degrees above the northern end of the horizon during the summer solstice, to directly overhead on the fall equinox, and then down to 66.5 degrees above the southern end of the horizon during the summer solstice (**Figure 6h-10**).

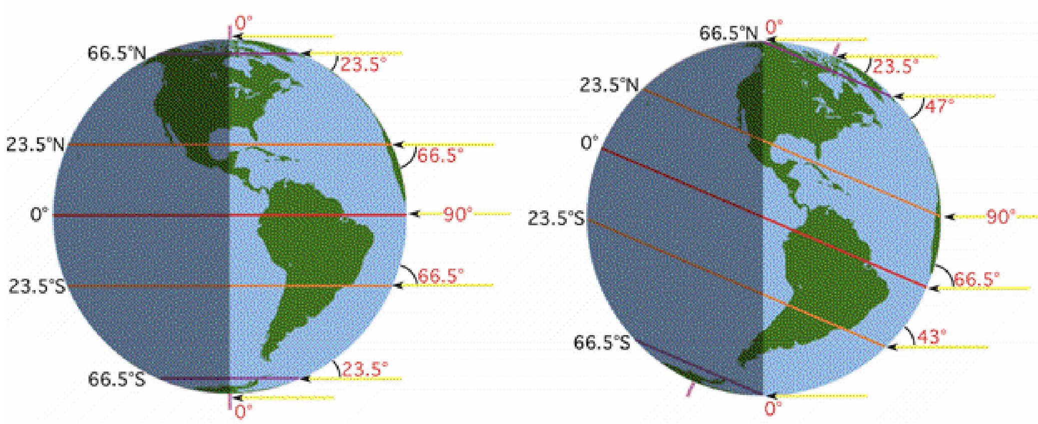


{PRIVATE}**Figure 6h-9:** Variations in solar altitude at solar noon for 50 degrees North during the summer solstice, equinox, and winter solstice.



{PRIVATE**}Figure 6h-10:** Variations in solar altitude at solar noon for the equator during the summer solstice, equinox, and winter solstice.

The location on the Earth where the sun is directly overhead at ***solar noon*** is known as the ***subsolar point***. The sunsolar point occurs on the equator during the equinoxes (**Figure 6h-11**). During the summer solstice, the subsolar point moves to the Tropic of Cancer because at this time the North Pole is tilted 23.5 degrees toward the sun. The subsolar point is located at the Tropic of Capricorn on the winter solstice. On this date, the South Pole is now tilted toward the sun (**Figures 6h-2** and **6h-3**).



{PRIVATE**}Figure 6h-11:** Relationship of maximum sun height to latitude for the equinox (left) and summer solstice (right). The red values on the right of the globes are maximum solar altitudes at ***solar noon***. Black numbers on the left indicate the location of the Equator, Tropic of Cancer (23.5 degrees N), Tropic of Capricorn (23.5 degrees S), Arctic Circle (66.5 degrees N), and the Antarctic Circle (66.5 degrees S). The location of the North and South Poles are also identified. During the equinox, the equator is the location on the Earth with a sun angle of 90 degrees for solar noon. Note how maximum sun height declines with latitude as you move away from the Equator. For each degree of latitude traveled maximum sun height decreases by the same amount. At equinox, you can also calculate the noon angle by substracting the location's latitude from 90. During the summer solstice, the sun is now directly overhead at the Tropic of Cancer. All locations above this location have maximum sun heights that are 23.5 degrees higher from the equinox situation. Places above the Arctic Circle are in 24 hours of daylight. Below the Tropic of Cancer the noon angle of the sun drops one degree in height for each degree of latitude traveled. At the Antarctic Circle, maximum sun height becomes 0 degrees and locations south of this point on the Earth are in 24 hours of darkness.

The following **table** describes the changes in solar altitude at solar noon for the two solstices and equinoxes. All measurements are in degrees (***horizon*** has 180 degrees from ***True North*** to ***True South***) and are measured from either True North or True South (whatever is closer).

{PRIVATE**}Table 6h-1:** Maximum sun altitudes for selected latitudes during the two solstices and equinoxes.

{PRIVATE**}Location's Latitude**	**Vernal Equinox March 21/22**	**Summer Solstice June 20/21**	**Autumnal Equinox September 22/23**	**Winter Solstice December 21/22**
90 N	**0 degrees**	**23.5 degrees**	**0 degrees**	**- 23.5 degrees**
70 N	**20 degrees**	**43.5 degrees**	**20 degrees**	**-3.5 degrees**
66.5 N	**23.5 degrees**	**47 degrees**	**23.5 degrees**	**0 degrees**
60 N	**30 degrees**	**53.5 degrees**	**30 degrees**	**6.5 degrees**
50 N	**40 degrees**	**63.5 degrees**	**40 degrees**	**16.5 degrees**
23.5 N	**66.5 degrees**	**90 degrees**	**66.5 degrees**	**43 degrees**
0 degrees	**90 degrees**	**66.5 degrees**	**90 degrees**	**66.5 degrees**
23.5 S	**66.5 degrees**	**43 degrees**	**66.5 degrees**	**90 degrees**
50 S	**40 degrees**	**16.5 degrees**	**40 degrees**	**63.5 degrees**
60 S	**30 degrees**	**6.5 degrees**	**30 degrees**	**53.5 degrees**
66.5 S	**23.5 degrees**	**0 degrees**	**23.5 degrees**	**47 degrees**
70 S	**20 degrees**	**-3.5 degrees**	**20 degrees**	**43.5 degrees**
90 S	**0 degrees**	**- 23.5 degrees**	**0 degrees**	**23.5 degrees**

The following links show graphical illustrations of the annual movements of the sun in our skies for selected latitudes. In these illustrations, solar angles are measured from both ***True North*** and ***True South*** for ***solar noon***.

| 90 N | 66.5 N | 50 N | 23.5 N | Equator (0) | 23.5 S | 50 S | 66.5 S | 90 S |

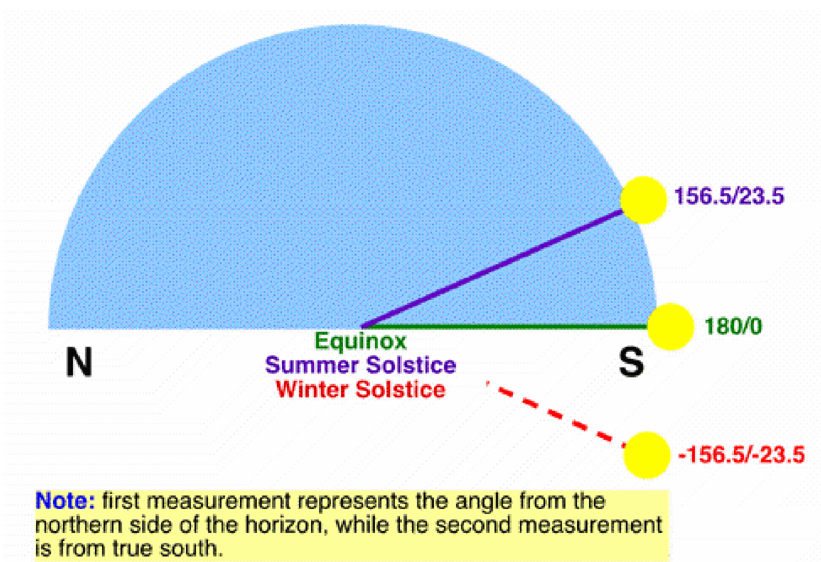


Figure 6h-8: Solar noon sun angles for 90 degrees N.

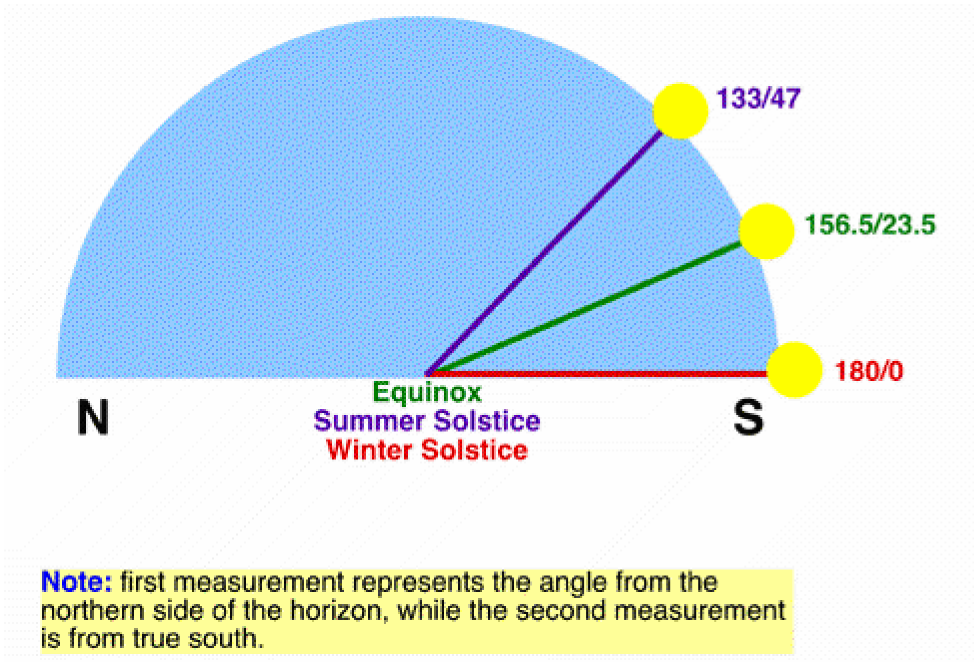


Figure 6h-9: Solar noon sun angles for 66.5 degrees N.

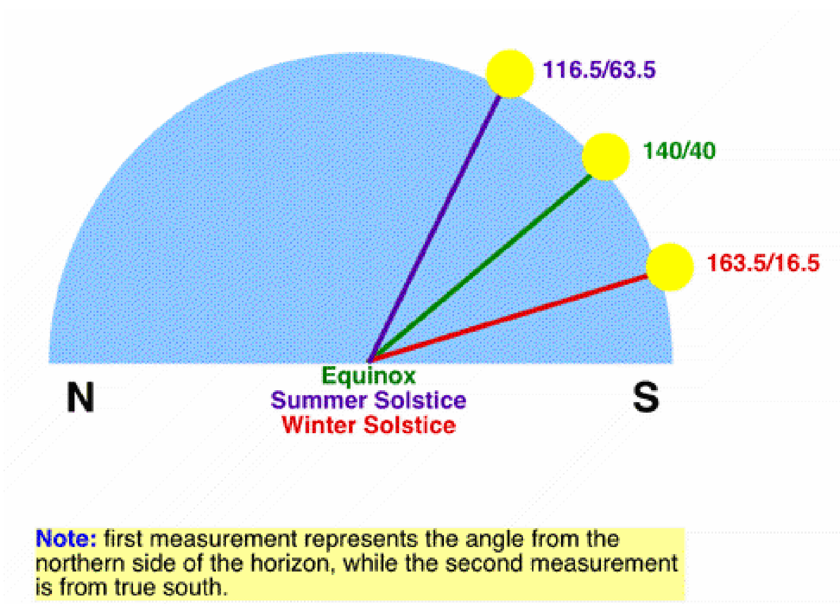


Figure 6h-10: Solar noon sun angles for 60 degrees N.

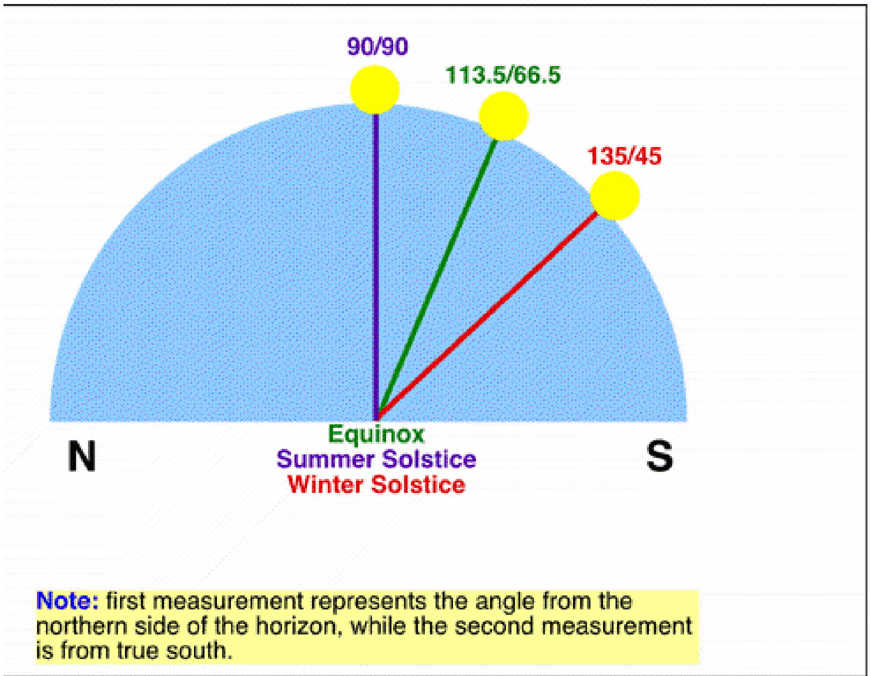


Figure 6h-11: Solar noon sun angles for 23.5 degrees N.

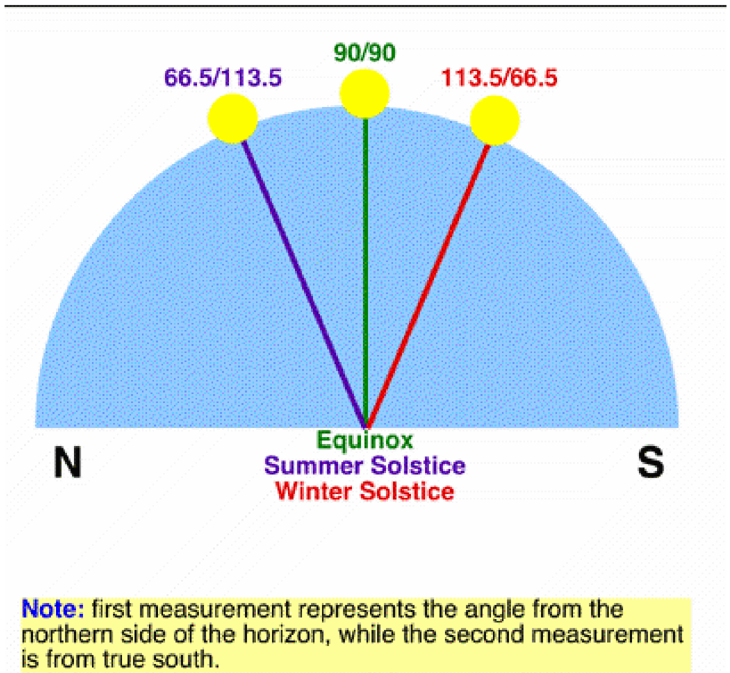


Figure 6h-12: Solar noon sun angles for the equator

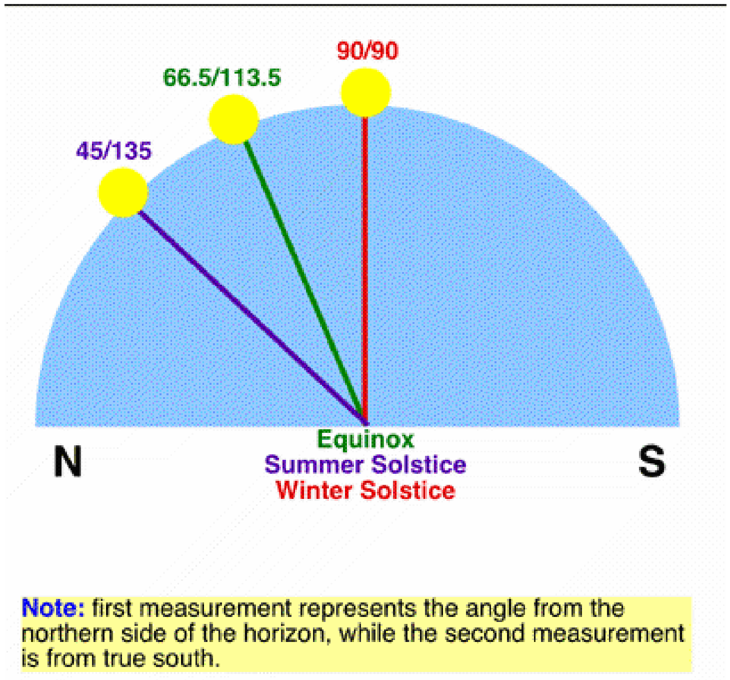


Figure 6h-13: Solar noon sun angles for 23.5 degrees S.

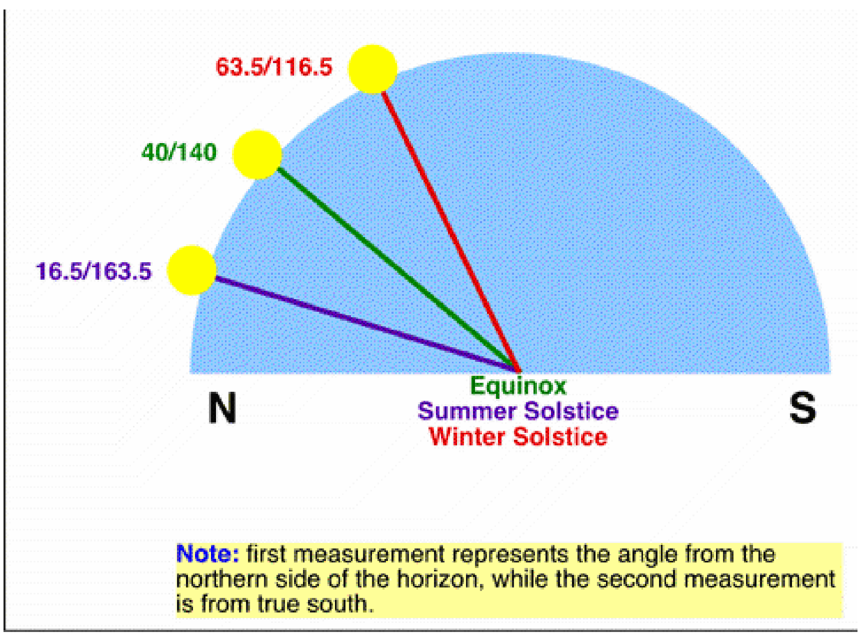


Figure 6h-14: Solar noon sun angles for 50 degrees S.

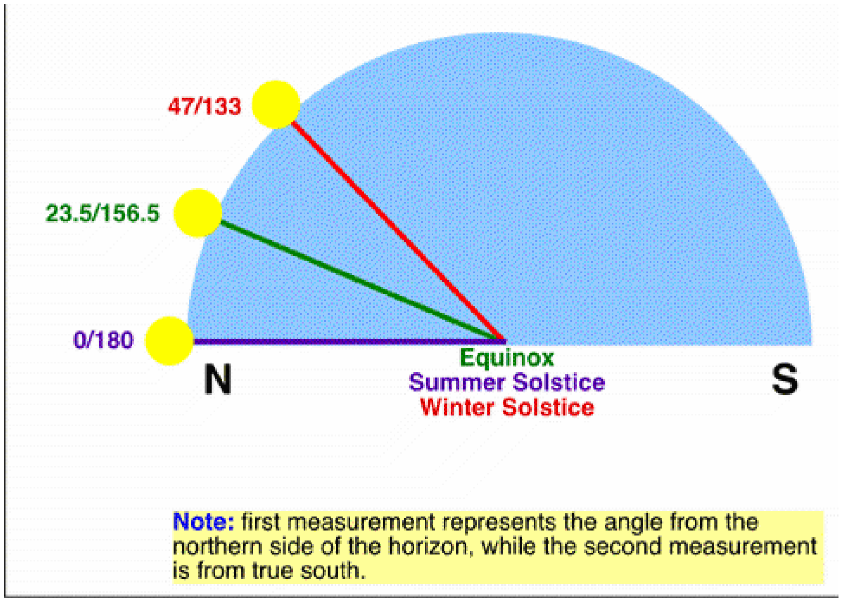


Figure 6h-15: Solar noon sun angles for 66.5 degrees S.

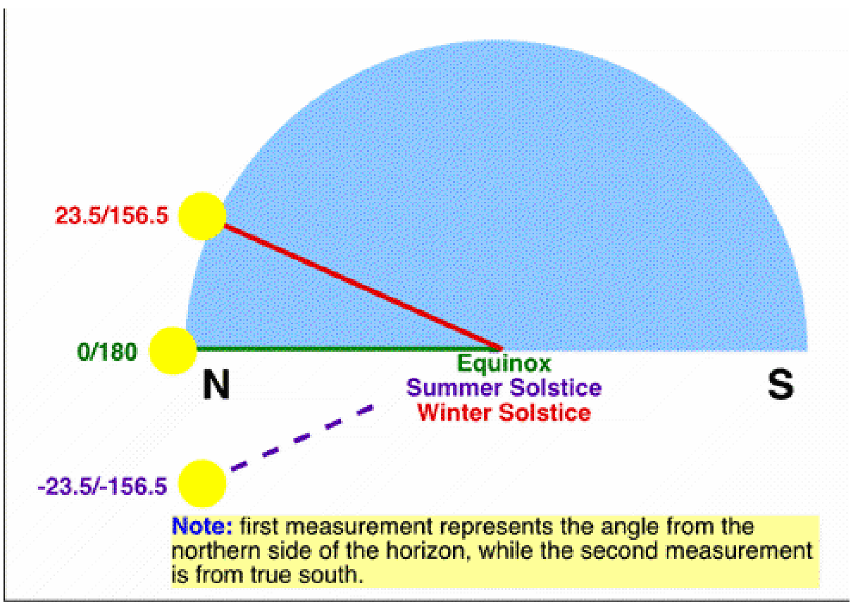


Figure 6h-16: Solar noon sun angles for 90 degrees S.

Finally, the altitude of the sun at solar noon can also be calculated with the following simple equation:

Altitude A = 90 - Latitude L +/- Declination D

In this equation, **L** is the latitude of the location in degrees and **D** is the declination. The equation is simplified to **A = 90 - L** if sun angle determinations are being made for either equinox date. If the sun angle determination is for a solstice date, declination (**D**) is added to latitude (**L**) if the location is experiencing summer (northern latitudes = summer solstice; southern latitudes = winter solstice) and subtracted from latitude (**L**) if the location is experiencing winter (northern latitudes = winter solstice; southern latitudes = summer solstice). All answers from this equation are given relative to ***True North*** for southern latitudes and ***True South*** for northern latitudes. For our purposes only the declinations of the two solstices and two equinoxes are important. These values are: Summer Solstice **D**=23.5, Winter Solstice **D**=23.5, Autumnal Equinox **D**=0, and Vernal Equinox **D**=0. When using the above equation in tropical latitudes, sun altitude values greater than 90 degrees may occur for some calculations. When this occurs, the noonday sun is actually behind you when looking towards the equator. Under these circumstances, sun altitude should be recalculated as follows:

Altitude A = 90 - (originally calculated Altitude A - 90)

(i) Earth-Sun Relationships and Insolation

{PRIVATE**}**In the previous topic, we learned that the Earth's seasons are controlled by changes in the **duration** and **intensity** of ***solar radiation*** or ***insolation***. Both of these factors are in turn governed by the annual change in the position of the Earth's axis relative to the sun (see ***Figure 6h-4***).

Yearly changes in the position of the Earth's axis cause the location of the sun to wander 47 degrees across our skies. Changes in the location of the sun have a direct effect on the intensity of solar radiation. The intensity of solar radiation is largely a function of the ***angle of incidence***, the angle at which the sun's rays strike the Earth's surface. If the sun is positioned directly overhead or 90 degrees from the horizon, the incoming insolation strikes the surface of the Earth at right angles and is most intense. If the sun is 45 degrees above the horizon, the incoming insolation strikes the Earth's surface at an angle. This causes the rays to be spread out over a larger surface area reducing the intensity of the radiation. **Figure 6i-1** models the effect of changing the angle of incidence from 90 to 45 degrees. As illustrated, the lower sun angle (45 degrees) causes the radiation to be received over a much larger surface area. This surface area is approximately 40 % greater than the area covered by an angle of 90 degrees. The lower angle also reduces the intensity of the incoming rays by 30 %.

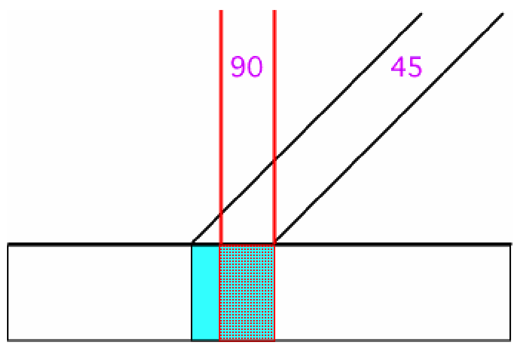


Figure 6i-1: Effect of angle on the area that intercepts an incoming beam of radiation.

We can also model the effect the angle of incidence has on insolation intensity with the following simple equation:

Intensity = SIN (A)

where, **A** is the angle of incidence and **SIN** is the sine function found on most calculators. Using this equation we can determine that an angle of 90 degrees gives us a value of 1.00 or 100 % (1.00 x 100). Let us compare this maximum value with values determined for other angles of incidence. **Note** the answers are expressed as a percentage of the potential maximum value.

SIN 80 = 0.98 or 98%

SIN 70 = 0.94 or 94%

SIN 60 = 0.87 or 87%

SIN 50 = 0.77 or 77%

SIN 40 = 0.64 or 64%

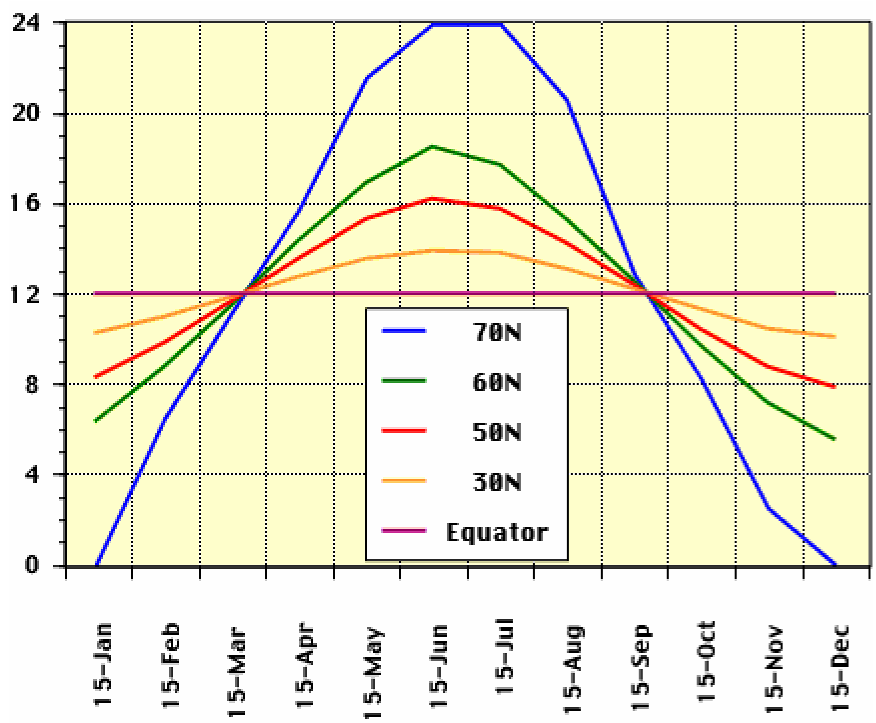
SIN 30 = 0.50 or 50%

SIN 20 = 0.34 or 34%

SIN 10 = 0.17 or 17%

SIN 0 = 0.00 or 0%

The yearly changes in the position of the Earth's axis relative to the ***plane of the ecliptic*** also causes seasonal variations in day length to all locations outside of the equator. Longest days occur during the ***summer solstice*** for locations north of the equator and on the ***winter solstice*** for locations in the Southern Hemisphere. The equator experiences equal day and night on every day of the year. Day and night is also of equal length for all Earth locations on the ***autumnal*** and ***vernal*** **equinoxes**. **Figure 6i-2** describes the change in the length of day for locations at the equator, 10, 30, 50, 60, and 70 degrees North over a one-year period. The illustration suggests that days are longer than nights in the Northern Hemisphere from the March equinox to the September equinox. Between the September to March equinox days are shorter than nights in the Northern Hemisphere. The opposite is true in the Southern Hemisphere. The graph also shows that the seasonal (winter to summer) variation in day length increases with increasing latitude.



{PRIVATE}**Figure 6i-2:** Annual variations in day length for locations at the equator, 30, 50, 60, and 70 degrees North latitude.

Figure 6i-3 below describes the potential insolation available for the equator and several locations in the Northern Hemisphere over a one-year period. The values plotted on this **graph** take into account the combined effects of angle of incidence and day length duration (see ***Table 6h-2***). Locations at the equator show the least amount of variation in insolation over a one-year period. These slight changes in insolation result only from the annual changes in the altitude of the sun above the horizon, as the duration of daylight at the equator is always 12 hours. The peaks in insolation intensity correspond to the two ***equinoxes*** when the sun is directly overhead. The two annual minimums of insolation occur on the ***solstices*** when the maximum height of the sun above the horizon reaches an angle of 66.5 degrees.

The most extreme variations in insolation received in the Northern Hemisphere occur at 90 degrees North. During the ***summer solstice*** this location receives more potential incoming solar radiation than any other location graphed. At this time the sun never sets. In fact, it remains at an altitude of 23.5 degrees above the horizon for the whole day. From September 22 (***autumnal equinox***) to March 21, (***vernal equinox***) no insolation is received at 90 degrees North. During this period the sun slips below the horizon as the northern axis of the Earth becomes tilted away from the sun.

The annual insolation curve for locations at 60 degrees North best approximates the seasonal changes in solar radiation intensity perceived at our latitude. Maximum values of insolation are received at the summer solstice when day length and angle of incidence are at their maximum (see ***Table 6h-2*** and **section *6h***). During the ***summer solstice*** day length is 18 hours and 27 minutes and the angle of the sun reaches a maximum value of 53.5 degrees above the horizon. Minimum values of insolation are received at the ***winter solstice*** when day length and angle of incidence are at their minimum (see ***Table 6h-2*** and **section *6h***). During the ***winter solstice*** day length is only 5 hours and 33 minutes and the angle of the sun reaches a lowest value of 6.5 degrees above the horizon.

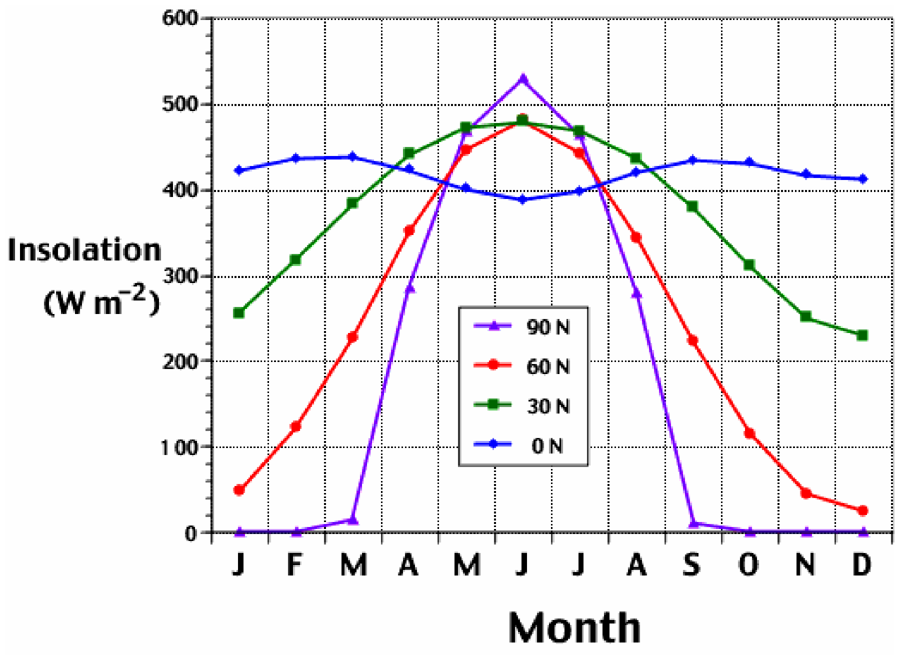


Figure 6i-3: Monthly values of available insolation for the equator, 30, 60, and 90 degrees North.

7) Introduction to Meteorology and Climatology

(a) Atmospheric Composition

{PRIVATE**}Table 7a-1** lists the eleven most abundant gases found in the Earth's lower atmosphere by volume. Of the gases listed, nitrogen, oxygen, water vapor, carbon dioxide, methane, nitrous oxide, and ozone are extremely important to the health of the Earth's biosphere.

The **table** indicates that **nitrogen** and **oxygen** are the main components of the atmosphere by volume. Together these two gases make up approximately 99 % of the dry atmosphere. Both of these gases have very important associations with life. Nitrogen is removed from the atmosphere and deposited at the Earth's surface mainly by specialized ***nitrogen fixing*** bacteria, and by way of ***lightning*** through precipitation. The addition of this nitrogen to the Earth's surface soils and various water bodies supplies much needed nutrition for plant growth. Nitrogen returns to the atmosphere primarily through biomass combustion and ***denitrification***.

Oxygen is exchanged between the atmosphere and life through the processes of ***photosynthesis*** and ***respiration***. Photosynthesis produces oxygen when carbon dioxide and water are chemically converted into glucose with the help of sunlight. Respiration is a process whose reciprocal is photosynthesis. In respiration, oxygen is combined with glucose to chemically release energy for metabolism. The products of this reaction are water and carbon dioxide.

The next most abundant gas on the **table** is **water vapor**. Water vapor varies in concentration in the atmosphere both spatially and temporally. The highest concentrations of water vapor are found near the equator over the oceans and tropical rain forests. Cold polar areas and subtropical continental deserts are locations where the volume of water vapor can approach zero percent. Water vapor has several very important functional roles on our planet:

It redistributes heat energy on the Earth through ***latent heat*** energy exchange.

The condensation of water vapor creates precipitaion that falls to the Earth's surface providing needed fresh water for plants and animals.

It helps warm the Earth's atmosphere through the ***greenhouse effect***.

The fifth most abundant gas in the atmosphere is ***carbon dioxide***. The volume of this gas has increased by over 25 % in the last three hundred years (see ***Figure 7a-1***). This increase is primarily due to human induced burning for fossil fuels, deforestation, and other forms of land-use change. Some scientists believe that this increase is causing ***global warming*** through an enhancement of the ***greenhouse effect***. Carbon dioxide is also exchanged between the atmosphere and life through the processes of ***photosynthesis*** and ***respiration***.

Methane is a very strong greenhouse gas. Since 1750, methane concentrations in the atmosphere have increased by more than 140 %. The primary sources for the additional methane added to the atmosphere (in order of importance) are: rice cultivation; domestic grazing animals; termites; landfills; coal mining; and, oil and gas extraction. Anaerobic conditions associated with rice paddy flooding results in the formation of methane gas. However, an accurate estimate of how much methane is being produced from rice paddies has been difficult to ascertain. More than 60 % of all rice paddies are found in India and China where scientific data concerning emission rates are unavailable. Nevertheless, scientists believe that the contribution of rice paddies is large because this form of crop production has more than doubled since 1950. Grazing animals release methane to the environment as a result of herbaceous digestion. Some researchers believe the addition of methane from this source has more than quadrupled over the last century. Termites also release methane through similar processes. Land-use change in the tropics, due to deforestation, ranching, and farming, may be causing termite numbers to expand. If this assumption is correct, the contribution from these insects may be important. Methane is also released from landfills, coal mines, and gas and oil drilling. Landfills produce methane as organic wastes decompose over time. Coal, oil, and natural gas deposits release methane to the atmosphere when these deposits are excavated or drilled.

The average concentration of the greenhouse gas ***nitrous oxide*** is now increasing at a rate of 0.2 to 0.3 % per year. Its part in the enhancement of the greenhouse effect is minor relative to the other greenhouse gases already mentioned. However, it does have an important role in the artificial fertilization of ecosystems. In extreme cases, this fertilization can lead to the death of forests, eutrophication of aquatic habitats, and species exclusion. Sources for the increase of nitrous oxide in the atmosphere include: land-use conversion; fossil fuel combustion; biomass burning; and soil fertilization. Most of the nitrous oxide added to the atmosphere each year comes from deforestation and the conversion of forest, savanna and grassland ecosystems into agricultural fields and rangeland. Both of these processes reduce the amount of nitrogen stored in living vegetation and soil through the decomposition of organic matter. Nitrous oxide is also released into the atmosphere when fossil fuels and biomass are burned. However, the combined contribution to the increase of this gas in the atmosphere is thought to be minor. The use of nitrate and ammonium fertilizers to enhance plant growth is another source of nitrous oxide. How much is released from this process has been difficult to quantify. Estimates suggest that the contribution from this source represents from 50 % to 0.2 % of nitrous oxide added to the atmosphere annually.

Ozone's role in the enhancement of the greenhouse effect has been difficult to determine. Accurate measurements of past long-term (more than 25 years in the past) levels of this gas in the atmosphere are currently unavailable. Moreover, concentrations of ozone gas are found in two different regions of the Earth's atmosphere. The majority of the ozone (about 97 %) found in the atmosphere is concentrated in the ***stratosphere*** at an altitude of 15 to 55 kilometers above the Earth's surface. This stratospheric ozone provides an important service to life on the Earth as it absorbs harmful ultraviolet radiation. In recent years, levels of **stratospheric ozone** have been decreasing due to the buildup of human created ***chlorofluorocarbons*** in the atmosphere. Since the late 1970s, scientists have noticed the development of severe holes in the ozone layer over Antarctica. Satellite measurements have indicated that the zone from 65 degrees North to 65 degrees South latitude has had a 3 % decrease in stratospheric ozone since 1978.

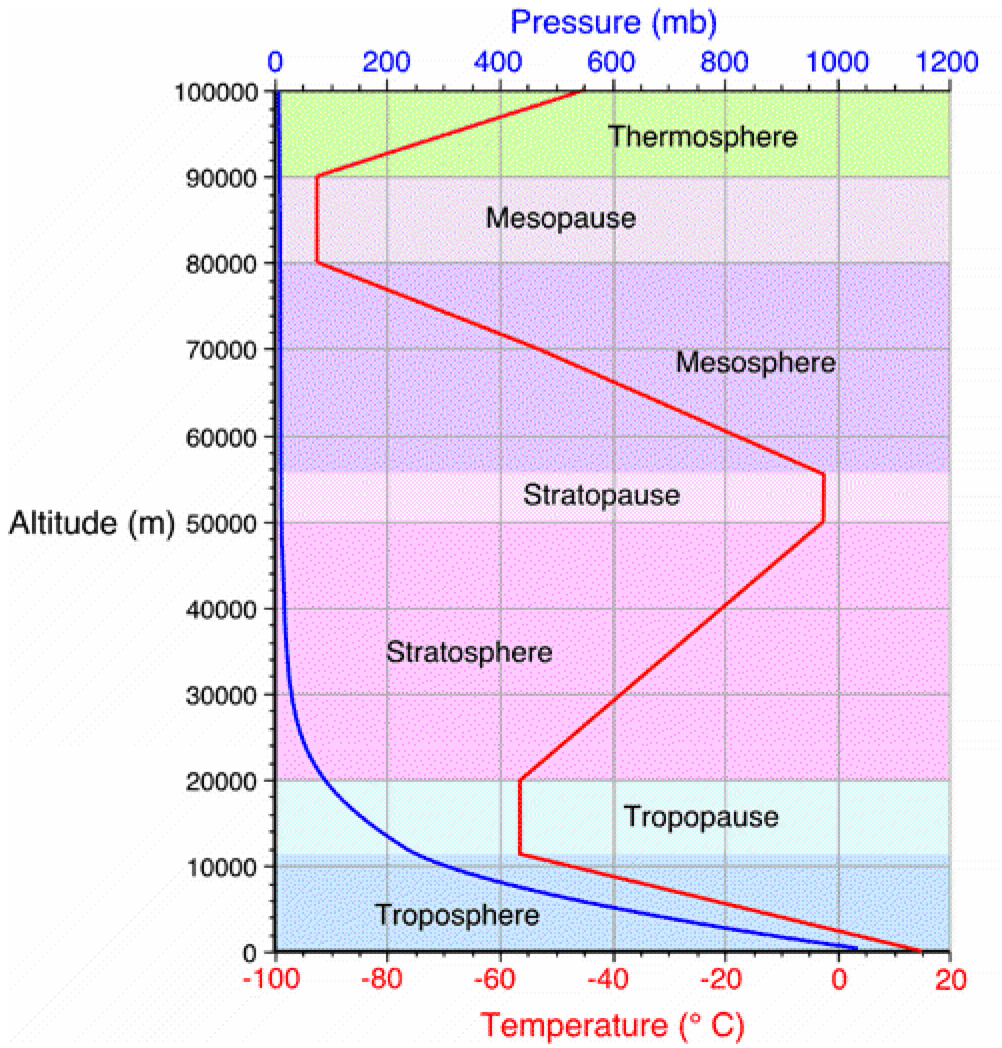
Ozone is also highly concentrated at the Earth's surface in and around cities. Most of this ozone is created as a by product of human created ***photochemical smog***. This buildup of ozone is toxic to organisms living at the Earth's surface.

Table 7a-1: Average composition of the atmosphere up to an altitude of 25 km.

{PRIVATE}Gas Name	Chemical Formula	Percent Volume
Nitrogen	N2	78.08%
Oxygen	O2	20.95%
*Water	H2O	0 to 4%
Argon	Ar	0.93%
*Carbon Dioxide	CO2	0.0360%
Neon	Ne	0.0018%
Helium	He	0.0005%
*Methane	CH4	0.00017%
Hydrogen	H2	0.00005%
*Nitrous Oxide	N2O	0.00003%
*Ozone	O3	0.000004%
*** variable gases**		

(b) The Layered Atmosphere

{PRIVATE}The Earth's atmosphere contains several different layers that can be defined according to air ***temperature*** or **chemical composition**. **Figure 7b-1** displays some of these layers in an average atmosphere.



{PRIVATE**}Figure 7b-1:** Vertical change in average global atmospheric temperature and pressure. Variations in the way temperature changes with height indicates the atmosphere is composed of a number of different layers (labeled above). These variations are due to alterations in the chemical and physical nature of the atmosphere with altitude.

According to temperature, the atmosphere contains seven different layers (**Figure 7b-1**). From the surface of the Earth to approximate 11 kilometers in altitude the layer called the ***troposphere*** exists. About 75 % of the total mass of the atmosphere is contained in this layer. It is also the layer where the majority of our weather occurs. Maximum air temperature also occurs near the Earth's surface in this layer. With increasing height, air temperature drops uniformly with altitude at a rate of approximately 6.5 degrees Celsius per 1000 meters. This phenomenon is commonly called the ***Environmental Lapse Rate***. At an average temperature of -55 degrees Celsius, the top of the troposphere is reached.

The ***tropopause***, extending from 11 to 20 kilometers, is an ***isothermal*** layer in the atmosphere where temperature remains constant over a distance of about 9 kilometers. It is also the layer in the atmosphere where the ***jet streams*** occur.

Above the tropopause, is the ***stratosphere***. This layer extends from an average altitude of 20 to 48 kilometers above the Earth's surface. In the stratosphere, temperature increases with altitude because a localized concentration of ***ozone*** gas molecules absorbs ultraviolet sunlight creating heat energy. Ozone is primarily found in the atmosphere at varying concentrations between the altitudes of 10 to 50 kilometers. This layer of ozone is also called the ***ozone layer*** . The ozone layer is important to organisms at the Earth's surface as it protects them from the harmful effects of the sun's ultraviolet radiation. Without the ozone layer life could not exist on the Earth's surface.

Separating the ***mesosphere*** from the stratosphere is another isothermal layer called the ***stratopause***. In the mesosphere, the atmosphere reaches its coldest temperatures (about -90 degrees Celsius) at a height of approximately 80 kilometers. Above the mesosphere is another isothermal layer called the ***mesopause***.

The last atmospheric layer, as defined by vertical temperature change, has an altitude greater than 90 kilometers and is called the ***thermosphere***. The thermosphere is the hottest layer in the atmosphere. Heat is generated from the absorption of solar radiation by oxygen molecules. Temperatures in this layer can reach 1300 to 1800 degrees Celsius.

(c) Physical Behavior of the Atmosphere and the Gas Laws

{PRIVATE**}**In the previous topic, we learned the atmosphere is composed of a mixture of many different gases. This mixture behaves in many ways as if it were a single gas. As a result of this phenomenon, the following generalizations describe important relationships between ***temperature***, ***pressure***, ***density*** and ***volume***, that relate to the Earth's atmosphere.

(1) When ***temperature*** is held constant, the ***density*** of a gas is ***proportional*** to ***pressure***, and ***volume*** is ***inversely proportional*** to ***pressure***. Accordingly, an increase in ***pressure*** will cause an increase in ***density*** of the gas and a decrease in its ***volume***.

(2) If ***volume*** is kept constant, the ***pressure*** of a unit mass of gas is ***proportional*** to ***temperature***. If ***temperature*** increase so will ***pressure***, assuming no change in the ***volume*** of the gas.

(3) Holding ***pressure*** constant, causes the ***temperature*** of a gas to be ***proportional*** to ***volume***, and ***inversely proportional*** to ***density***. Thus, increasing ***temperature*** of a unit mass of gas causes its ***volume*** to expand and its ***density*** to decrease as long as there is no change in ***pressure***.

These relationships can also be described mathematically by the ***Ideal Gas Law***. Two equations that are commonly used to describe this law are:

Pressure x Volume = Constant x Temperature

and

Pressure = Density x Constant x Temperature

(d) Atmospheric Pressure

{PRIVATE**}Introduction**

Air is a tangible material substance and as a result has ***mass***. Any object with mass is influenced by the universal force known as ***gravity***. Newton's **Law of Universal Gravitation** states: any two objects separated in space are attracted to each other by a force proportional to the product of their masses and inversely proportional to the square of the distance between them. On the Earth, gravity can also be expressed as a ***force of acceleration*** of about 9.8 meters per second per second. As a result of this force, any object falling towards the surface of the Earth accelerates at an accelerating rate (1st second - 9.8 meters per second, 2nd second - 19.6 meters per second, 3rd second - 29.4 meters per second, and so on.) until ***terminal velocity*** is attained.

Gravity shapes and influences all atmospheric processes. It causes the ***density*** and ***pressure*** of air to decrease exponentially as one moves away from the surface of the Earth. **Figure 7d-1** below models the average change in air pressure with height above the Earth's surface. In this graph, air pressure at the surface is illustrated as being approximately 1013 ***millibars*** (**mb**) or 1 kilogram per square centimeter of surface area.

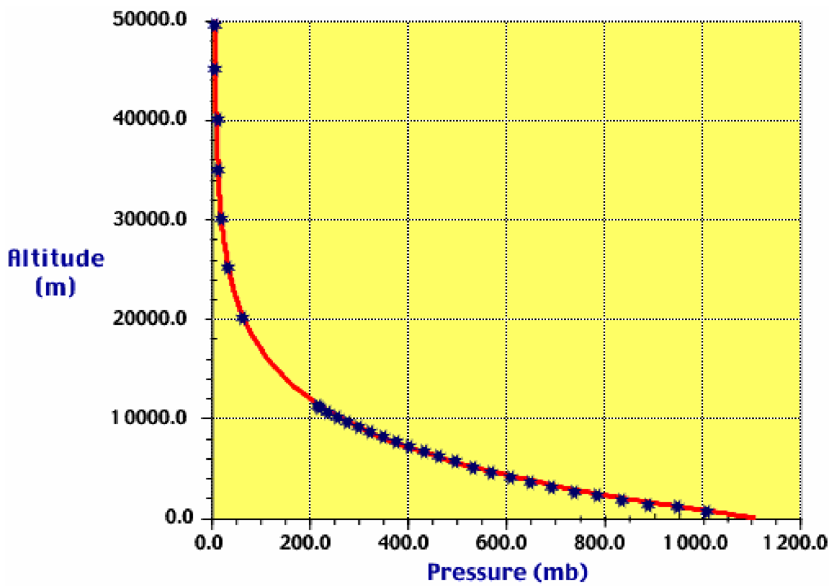


Figure 7d-1: Change in average atmospheric pressure with altitude.

Measuring Atmospheric Pressure

Any instrument that measures air pressure is called a **barometer**. The first measurement of atmospheric pressure began with a simple experiment performed by **Evangelista Torricelli** in 1643. In his experiment, Torricelli immersed a tube, sealed at one end, into a container of mercury (see **Figure 7d-2** below). Atmospheric pressure then forced the mercury up into the tube to a level that was considerably higher than the mercury in the container. Torricelli determined from this experiment that the pressure of the atmosphere is approximately 30 inches or 76 centimeters (one centimeter of mercury is equal to 13.3 ***millibars***). He also noticed that height of the mercury varied with changes in outside weather conditions.

Torricelli's Barometer

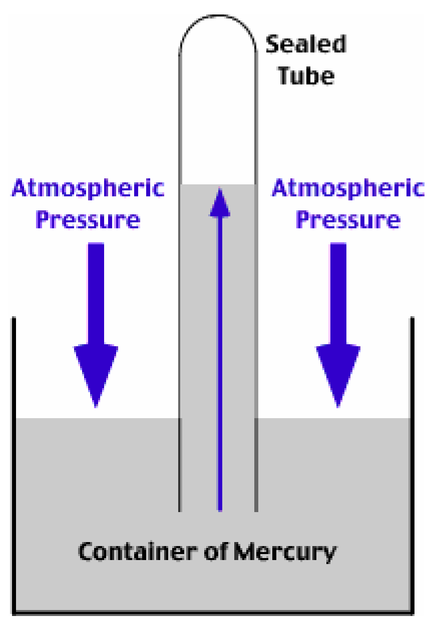


Figure 7d-2: Diagram showing the construction of Torricelli's barometer.

The most common type barometer used in homes is the **aneroid barometer** (**Figure 7d-3**). Inside this instrument is a small, flexible metal capsule called an aneroid cell. In the construction of the device, a vacuum is created inside the capsule so that small changes in outside air pressure cause the capsule to expand or contract. The size of the aneroid cell is then calibrated and any change in its volume is transmitted by springs and levers to an indicating arm that points to the corresponding atmospheric pressure.

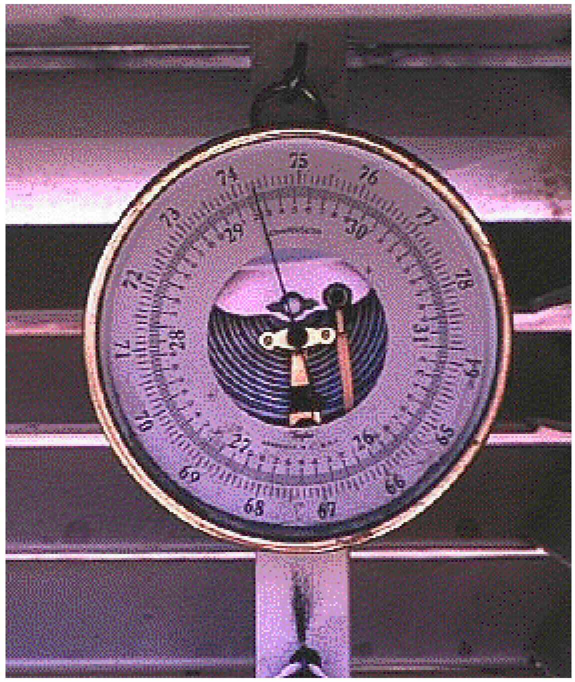


Figure 7d-3: Aneroid barometer.

For climatological and meteorological purposes, ***standard sea-level pressure*** is said to be 76.0 cm or 29.92 inches or 1013.2 ***millibars***. Scientists often use the ***kilopascal*** (**kPa**) as their preferred unit for measuring pressure. 1 kilopascal is equal to 10 millibars.

Atmospheric Pressure at the Earth's Surface

Figure 7d-4 describes monthly average sea-level pressure for the Earth's surface. This animation indicates that surface air pressure varies both spatially and temporally. During the winter months (December to February), areas of high pressure develop over central Asia (**Siberian High**), off the coast California (**Hawaiian High**), central North America (**Canadian High**), over Spain and northwest Africa extending into the subtropical North Atlantic (**Azores High**), and over the oceans in the Southern Hemisphere at the subtropics. Areas of low pressure occur just south of the Aleutian Islands (**Aleutian Low**), at the southern tip of Greenland (**Iceland Low**), and latitudes 50 to 80 degrees South.

During the summer months (June to August), a number of dominant winter pressure systems disappear. Gone are the **Siberian High** over central Asia and the dominant low pressure systems near the Aleutian Islands and at the southern tip of Greenland. The **Hawaiian** and **Azores High** intensify and expand northward into their relative ocean basins. High pressure systems over the subtropical oceans in Southern Hemisphere also intensity and expand to the north. New areas of dominant high pressure develop over Australia and Antarctica (**South Polar High**). Regions of low pressure form over central Asia and southwest Asia (**Asiatic Low**). These pressure systems are responsible for the summer ***monsoon*** rains of Asia.

We will examine this graphic again in **topic *7p*** when global circulation is discussed.

{PRIVATE}**Figure 7d-4:** Monthly average sea-level pressure and prevailing winds for the Earth's surface, 1959-1997. Atmosphere pressure values are adjusted for elevation and are described relative to sea -level. The slider at the bottom of the image allows you change the time of month. Pressure values are indicated by color shading. Blue shades indicate pressure lower than the global average, while yellow to orange shades are higher than average measurements. (**Source:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

(To view this animation your browser must have Apple's **QuickTime** plug-in and your monitor should be set for thousands of colors. The **QuickTime** plug-in is available for Macintosh, Windows 95, Windows 98 and Windows NT computers and can be downloaded from the World Wide Web site ***www.apple.com/quicktime***).

(e) The Ozone Layer

{PRIVATE}The ***ozone layer*** is a region of concentration of the ***ozone*** (O3) molecule in the Earth's atmosphere. The layer sits at an altitude of about 10 to 50 kilometers, with a maximum concentration in the ***stratosphere*** at an altitude of approximately 25 kilometers. In recent years, scientists have measured a seasonal thinning of the ozone layer primarily at the South Pole. This phenomenon is being called the ***ozone hole***.

The ozone layer naturally shields Earth's life from the harmful effects of the sun's ***ultraviolet (UV) radiation***. A severe decrease in the concentration of ozone in the ozone layer could lead to the following harmful effects:

An increase in the incidence of ***skin cancer*** (ultraviolet radiation can destroy acids in ***DNA***).

A large increase in cataracts and sun burning.

Suppression of immune systems in organisms.

Adverse impact on crops and animals.

Reduction in the growth of phytoplankton found in the Earth's oceans.

Cooling of the Earth's stratosphere and possibly some surface climatic effect.

Ozone is created naturally in the stratosphere by the combining of **atomic oxygen** (O) with **molecular oxygen** (O2). This process is activated by sunlight. Ozone is destroyed naturally by the absorption of ultraviolet radiation,

O3 + UV = O2 + O,

and by the collision of ozone with other atmospheric atoms and molecules.

O3 + O = 2O2

O3 + O3 = 3O2

Since the late 1970s, scientists have discovered that stratospheric ozone amounts over Antarctica in springtime (**September - November**) have decreased by as much as 60 %. Satellite measurements (**NIMBUS 7 - Total Ozone Mapping Spectrometer**) have indicated a 3 % decrease in ozone between 65 degrees North - 65 degrees South since 1978. A reduction of about 3 % per year has been measured for Antarctica where most of the ozone loss is occurring globally. During the late 1990s, large losses of ozone were recorded above Antarctica year after year in the months of September and August. In some years, spring levels of stratospheric ozone were more than 50 % lower than the levels recorded months prior to the seasonal development of the hole. The following animation describes the change in ozone levels at the South Pole during the period 1978 to 1992:

Figure 7e-1: Quicktime Animation of Antarctic Ozone Levels 1978-1992.

The following two animations describe 1979 and 1992 ozone levels at the South Pole for the period September 15 to November 15. View these images for comparison:

Figure 7e-2: Quicktime Animation of Antarctic Ozone Levels 1979; and

Figure 7e-3: Quicktime Animation of Antarctic Ozone Levels 1992.

It appears that human activities are altering the amount of stratospheric O3. The main agent responsible for this destruction was human-made ***chlorofluorocarbons*** or ***CFCs***. First produced by **General Motors Corporation** in 1928, CFCs were created as a replacement to the toxic refrigerant ***ammonia***. CFCs have also been used as a propellant in spray cans, cleaner for electronics, sterilant for hospital equipment, and to produce the bubbles in styrofoam. CFCs are cheap to produce and are very stable compounds, lasting up to 200 years in the atmosphere. By 1988, some 320,000 metric tons of CFCs were used worldwide.

In 1987, a number of nations around the world met to begin formulating a global plan, known as the ***Montreal Protocol***, to reduce and eliminate the use of CFCs. Since 1987, the plan has been amended in 1990 and 1992 to quicken the schedule of production and consumption reductions. By 1996, 161 countries were participating in the Protocol. The Montreal Protocol called for a 100 % reduction in the creation and use of CFCs by January 1, 1996 in the world's more developed countries. Less developed countries have until January 1, 2010 to stop their production and consumption of these dangerous chemicals.

CFCs created at the Earth's surface drift slowly upward to the stratosphere where ultraviolet radiation from the sun causes their decomposition and the release of **chlorine** (Cl). Chlorine in turn attacks the molecules of ozone chemically converting them into oxygen molecules.

Cl + O3 = ClO + O2

ClO + O = Cl + O2

A single chlorine atom removes about 100,000 ozone molecules before it is taken out of operation by other substances. Chlorine is removed from the stratosphere by two chemical reactions:

ClO + NO2 = ClONO2

CH4 + Cl = HCl + CH3

Normally, these two reactions would quickly neutralize the chlorine released into the stratosphere. However, the presence of ***polar stratospheric clouds***, rich in **nitrogen**, and sunlight facilitates a series of reactions which prolongs the reactive life of chlorine in the atmosphere. Interestingly, these polar stratospheric clouds require very cold air (approximately -85 degrees Celsius) for their formation. Stratospheric air of this temperature occurs normally every year above Antarctica in the winter and early spring months. Destruction of the ozone begins in Antarctica in the spring as this region moves from 24 hours of night to 24 hours of day. These clouds are less frequent in the Arctic stratosphere because winter cooling of the air in the stratosphere is less severe.

NASA's ***Earth Probe -Total Ozone Mapping Spectrometer home page*** has the latest images describing the current status of global stratosphere ozone levels in the atmosphere.

(f) Atmospheric Effects on Incoming Solar Radiation

{PRIVATE**}**Three atmospheric processes modify the solar radiation passing through our atmosphere destined to the Earth's surface. These processes act on the radiation when it interacts with gases and suspended particles found in the atmosphere. The process of ***scattering*** occurs when small particles and gas molecules ***diffuse*** part of the incoming solar radiation in random directions without any alteration to the wavelength of the electromagnetic energy (**Figure 7f-1**). Scattering does, however, reduce the amount of incoming radiation reaching the Earth's surface. A significant proportion of scattered shortwave solar radiation is redirected back to space. The amount of scattering that takes place is dependent on two factors: wavelength of the incoming radiation and the size of the scattering particle or gas molecule. In the Earth's atmosphere, the presence of a large number of particles with a size of about 0.5 microns results in shorter wavelengths being preferentially scattered. This factor also causes our sky to look blue because this color corresponds to those wavelengths that are best diffused. If scattering did not occur in our atmosphere the daylight sky would be black.

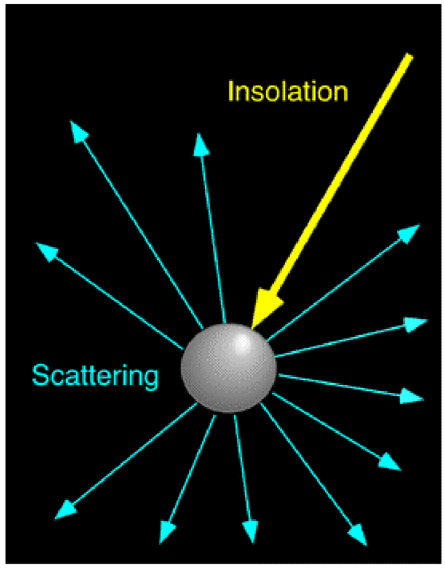


Figure 7f-1: Atmospheric scattering.

If intercepted, some gases and particles in the atmosphere have the ability to **absorb** incoming insolation (**Figure 7f-2**). ***Absorption*** is defined as a process in which solar radiation is retained by a substance and converted into ***heat energy***. The creation of heat energy also causes the substance to emit its own radiation. In general, the absorption of solar radiation by substances in the Earth's atmosphere results in temperatures that get no higher than 1800 degrees Celsius. According to ***Wien's Law***, bodies with temperatures at this level or lower would emit their radiation in the longwave band. Further, this emission of radiation is in all directions so a sizable proportion of this energy is lost to space.

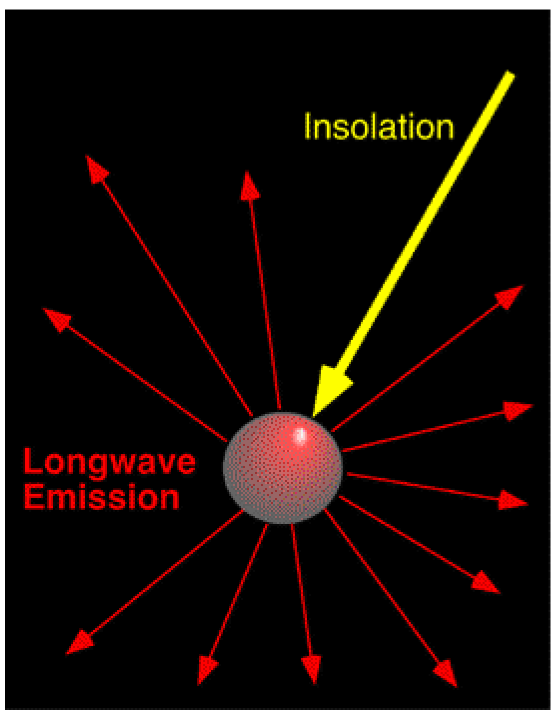


Figure 7f-2: Atmospheric absorption.

The final process in the atmosphere that modifies incoming solar radiation is ***reflection*** (**Figure 7f-3**). Reflection is a process where sunlight is redirect by 180 degrees after it strikes an atmospheric particle. This redirection causes a 100 % loss of the insolation. Most of the reflection in our atmosphere occurs in clouds when light is intercepted by particles of liquid and frozen water. The reflectivity of a cloud can range from 40 to 90 %.

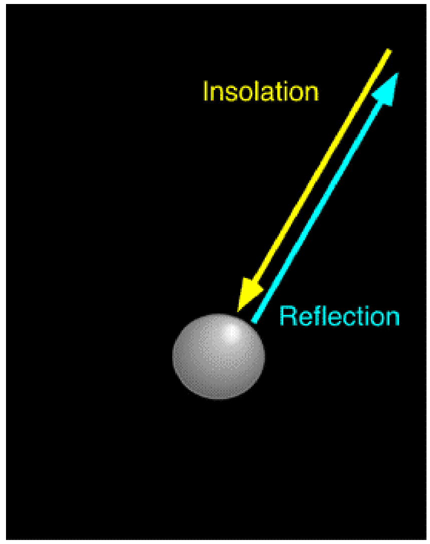


Figure 7f-3: Atmospheric reflection.

Sunlight reaching the Earth's surface unmodified by any of the above atmospheric processes is termed ***direct solar radiation***. Solar radiation that reaches the Earth's surface after it was altered by the process of scattering is called ***diffused solar radiation***. Not all of the direct and diffused radiation available at the Earth's surface is used to do **work** (photosynthesis, creation of sensible heat, evaporation, etc.). As in the atmosphere, some of the radiation received at the Earth's surface is redirected back to space by reflection. The following **image** describes the spatial pattern of surface ***reflectivity*** as measured for the year 1987.

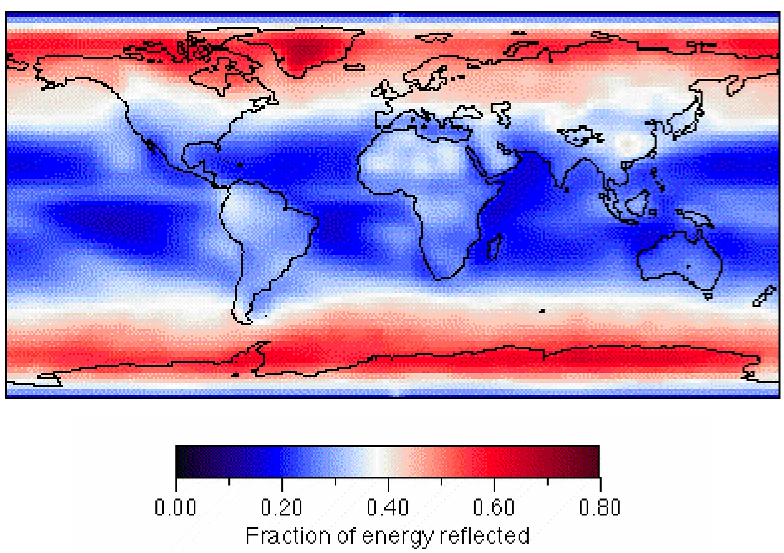


Figure 7f-4: Annual (1987) reflectivity of the Earth's surface.

(Image produced by the *CoVis Greenhouse Effect Visualizer*).

The ***reflectivity*** or ***albedo*** of the Earth's surface varies with the type of material that covers it. For example, fresh snow can reflect up to 95 % of the insolation that reaches it surface. Some other surface type reflectivities are:

Dry sand 35 to 45 %

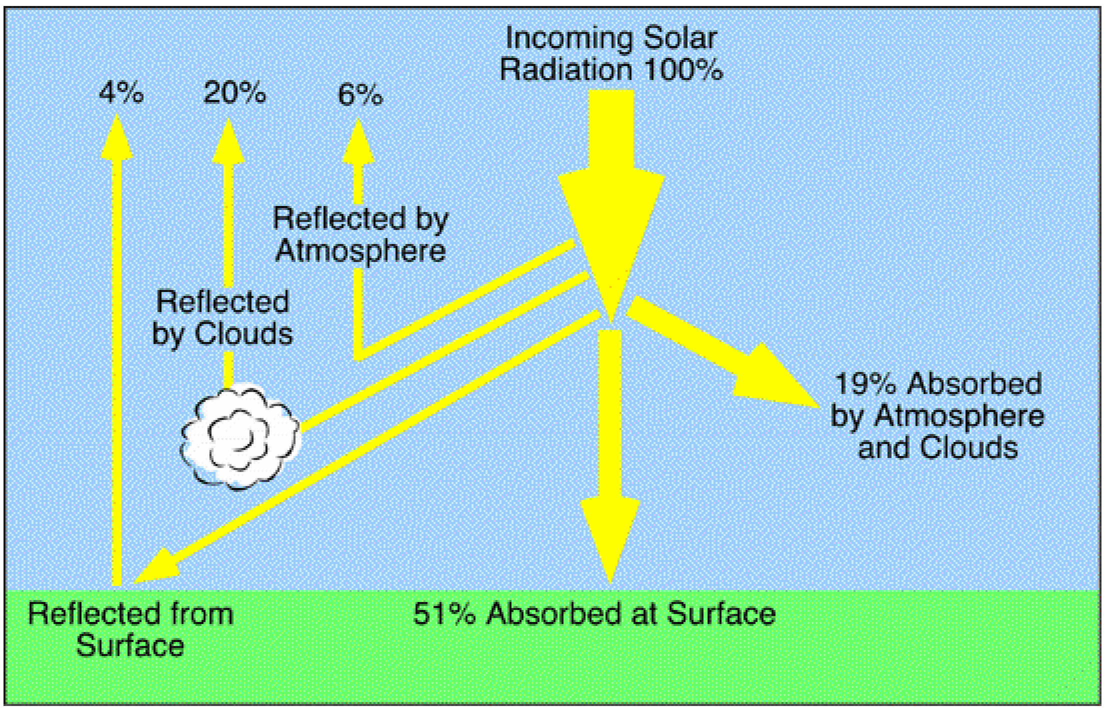
Broadleaf deciduous forest 5 to 10 %

Needleleaf confierous forest 10 to 20 %

Grass type vegetation 15 to 25 %

Reflectivity of the surface is often described by the term **surface *albedo***. The Earth's average **albedo**, reflectance from both the atmosphere and the surface, is about 30 %.

Figure 7f-5 describes the modification of solar radiation by atmospheric and surface processes for the whole Earth over a period of one year. Of all the sunlight that passes through the atmosphere annually, only 51 % is available at the Earth's surface to do work. This energy is used to ***heat*** the Earth's surface and lower atmosphere, **melt** and ***evaporate*** water, and run ***photosynthesis*** in plants. Of the other 49 %, 4 % is ***reflected*** back to space by the Earth's surface, 26 % is ***scattered*** or ***reflected*** to space by clouds and atmospheric particles, and 19 % is ***absorbed*** by atmospheric gases, particles, and clouds.



{PRIVATE}**Figure 7f-5:** Global modification of incoming solar radiation by atmospheric and surface processes.

(g) Global Patterns of Insolation Receipts

{PRIVATE}The following **image** describes the annual pattern of ***solar radiation absorption*** at the Earth's surface for the year 1987.

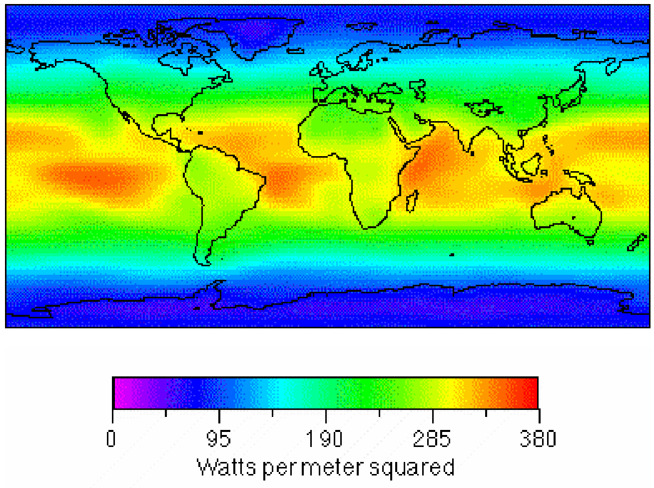


Figure 7g-1: Annual (1987) pattern of solar radiation absorbed at the Earth's surface.

(Image created by the *CoVis Greenhouse Effect Visualizer*).

The combined effect of Earth-sun relationships (***angle of incidence*** and day length variations) and the modification of the solar beam as it passes through the atmosphere produces specific global patterns of annual ***insolation*** receipt as seen on **Figure 7g-1** above (and see the **NASA WWW links** below). After examining these patterns, the following trends can be identified:

Highest values of insolation received occur in tropical latitudes. Within this zone there are localized maximums over the tropical oceans and deserts where the atmosphere has virtually no cloud development for most of the year. Insolation quantities at the equator over land during the solstices are approximately the same as values found in the middle latitudes during their summer (see **NASA links** below).

Outside the tropics, annual receipts of solar radiation generally decrease with increasing latitude. Minimum values occur at the poles. This pattern is primarily the result of Earth-sun geometric relationships and its effect on the duration and intensity of solar radiation received.

In middle and high latitudes, insolation values over the ocean, as compared to those at the same latitude over the land, are generally higher (see **NASA** images). Greater cloudiness over land surfaces accounts for this variation.

NASA's *Surface Radiation Budget Project* has used satellite data, computer models, and meteorological data to determine shortwave surface radiation fluxes for the period July 1983 to June 1991. The following **links** display these fluxes for January and July globally:

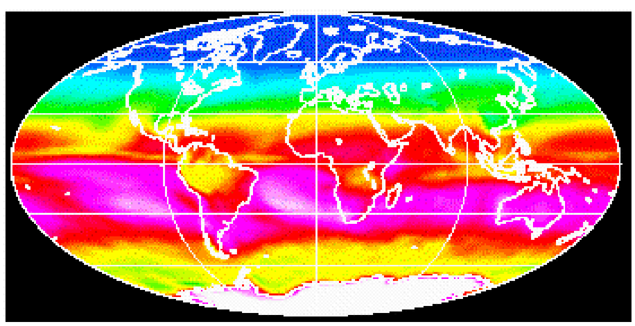


Figure 7g-2: Average available solar insolation at the Earth's surface: January 1984-1991. Highest values of available solar insolation occur at the South Pole due to high solar input and little cloud cover. High values also occur along the subtropical oceans of the Southern Hemisphere. Color range: blue - red - white, Values: 0 - 350W/m**2. Global mean = 187W/m**2, Minimum = 0W/m**2, Maximum = 426W/m**2. (**Source:** NASA *Surface Radiation Budget Project*). **(K + k)**

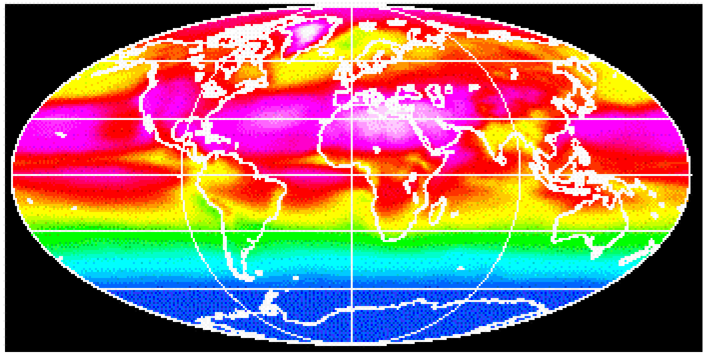


Figure 7g-2: Average available solar insolation at the Earth's surface: July 1983-1990. Highest values of available solar insolation over Greenland due to high solar input and low cloud amounts. High values also occur along the subtropics of the Northern Hemisphere. Color range: blue - red - white, Values: 0 - 350W/m**2. Global mean = 180W/m**2, Minimum = 0W/m**2, Maximum = 351/m**2. (**Source:** NASA *Surface Radiation Budget Project*). **(K + k)**

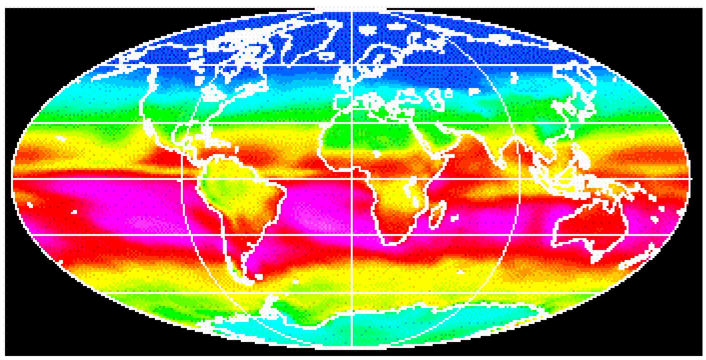


Figure 7g-4: Average absorbed solar insolation at the Earth's surface: January 1984-1991. Highest values occur along the subtropical oceans of the Southern Hemisphere. Lowest values occur over areas of high surface reflection such as the South Pole, cloudy regions, and areas of low solar input like the high latitudes of the Northern Hemisphere. Color range: blue - red - white, Values: 0 - 350W/m**2. Global mean = 162W/m**2, Minimum = 0W/m**2, Maximum = 315W/m**2. (**Source:** NASA *Surface Radiation Budget Project*).

[(K + k)(1 - a)]

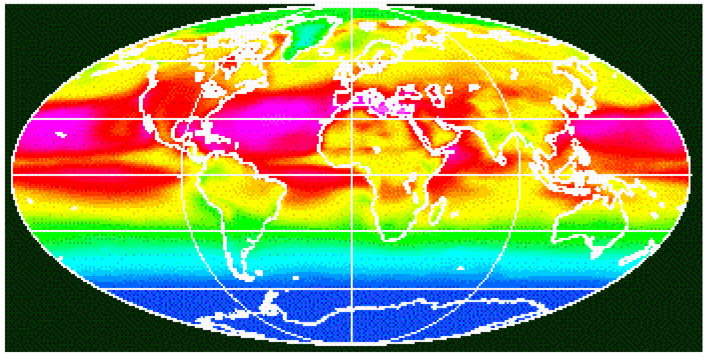


Figure 7g-5: Average absorbed solar insolation at the Earth's surface: July 1983-1990. Highest values occur over the subtropical oceans of the Northern Hemisphere due to high solar input and little cloud coverage. Lowest values occur in areas of high surface reflection such as snow/ice covered surfaces like Greenland, cloudy regions such as storm tracks, and areas of low solar input like the high latitudes of the Southern Hemisphere. Color range: blue - red - white, Values: 0 - 350W/m**2. Global mean = 158/m**2, Minimum = 0W/m**2, Maximum = 323/m**2. (**Source:** NASA *Surface Radiation Budget Project*). **[(K + k)(1 - a)]**

In the equations above, the mathematical terms have the following definitions (see ***topic 7(i)*** for more information on **radiation balance equations**):

K = Shortwave Direct Radiation

k = Shortwave Indirect Radiation

a = Reflectivity of the Surface or Surface Albedo

(h) The Greenhouse Effect

{PRIVATE**}**The ***greenhouse effect*** is a naturally occurring process that aids in heating the Earth's surface and atmosphere. It results from the fact that certain atmospheric gases, such as ***carbon dioxide***, **water vapor**, and ***methane***, are able to change the energy balance of the planet by absorbing ***longwave radiation*** emitted from the Earth's surface. Without the greenhouse effect life on this planet would probably not exist as the average temperature of the Earth would be a chilly -18 degrees Celsius, rather than the present 15 degrees Celsius.

As energy from the sun passes through the atmosphere a number of things take place (see ***Figure 7h-1***). A portion of the energy (26 % globally) is ***reflected*** back to space by clouds and particles. About 19 % of the energy available is absorbed by clouds, gases (like ***ozone***), and particles in the atmosphere. Of the remaining 55 % of the solar energy passing through the Earth's atmosphere, 4 % is reflected from the surface back to space. On average, about 51 % of the sun's radiation reaches the surface. This energy is then used in a number of processes, including the heating of the ground surface; the melting of ice and snow and the evaporation of water; and plant ***photosynthesis***.

The heating of the ground by sunlight causes the Earth's surface to become a radiator of energy in the longwave band (sometimes called ***infrared radiation***). This emission of energy is generally directed to space (see ***Figure 7h-2***). However, only a small portion of this energy actually makes it back to space. The majority of the outgoing infrared radiation is absorbed by the ***greenhouse gases*** (see **Figure 7h-3** below).

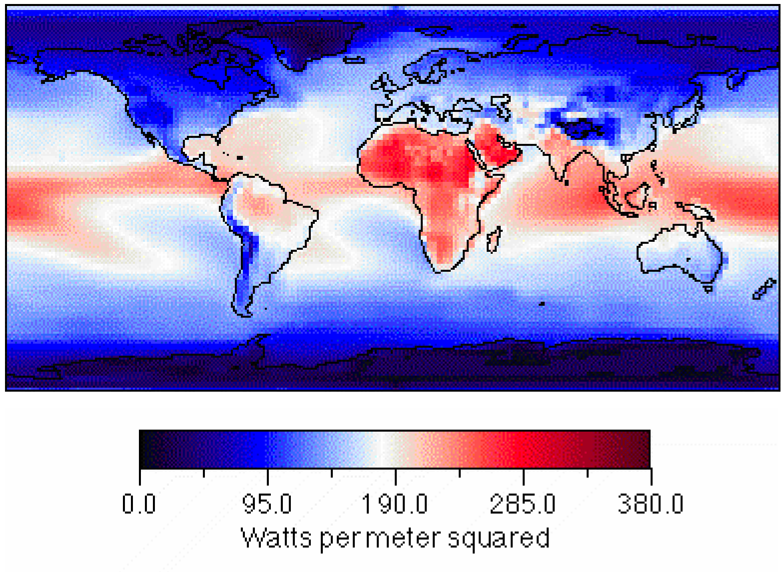


Figure 7h-3: Annual (1987) quantity of outgoing longwave radiation absorbed in the atmosphere.

(Image created by the *CoVis Greenhouse Effect Visualizer*).

Absorption of longwave radiation by the atmosphere causes additional heat energy to be added to the Earth's atmospheric system. The now warmer atmospheric greenhouse gas molecules begin radiating longwave energy in all directions. Over 90 % of this emission of longwave energy is directed back to the Earth's surface where it once again is absorbed by the surface. The heating of the ground by the longwave radiation causes the ground surface to once again radiate, repeating the cycle described above, again and again, until no more longwave is available for absorption.

The amount of heat energy added to the atmosphere by the ***greenhouse effect*** is controlled by the concentration of greenhouse gases in the Earth's atmosphere. All of the major greenhouse gases have increased in concentration since the beginning of the ***Industrial Revolution*** (about 1700 AD). As a result of these higher concentrations, scientists predict that the greenhouse effect will be **enhanced** and the Earth's climate will become warmer. Predicting the amount of warming is accomplished by computer modeling. Computer models suggest that a doubling of the concentration of the main greenhouse gas, ***carbon dioxide***, may raise the average global temperature between 1 and 3 degrees Celsius. However, the numeric equations of computer models do not accurately simulate the effects of a number of possible ***negative feedbacks***. For example, many of the models cannot properly simulate the negative effects that increased cloud cover would have on the radiation balance of a warmer Earth. Increasing the Earth's temperature would cause the oceans to evaporate greater amounts of water, causing the atmosphere to become cloudier. These extra clouds would then reflect a greater proportion of the sun's energy back to space reducing the amount of solar radiation absorbed by the atmosphere and the Earth's surface. With less solar energy being absorbed at the surface, the effects of an enhanced greenhouse effect may be counteracted.

A number of gases are involved in the greenhouse effect (see ***Table 7h-1*** below). These gases include: ***carbon dioxide*** (CO_2); ***methane*** (CH_4); ***nitrous oxide*** (N_2O); ***chlorofluorocarbons*** (CFxClx); and ***tropospheric ozone*** (O_3). Of these gases, the single most important gas is ***carbon dioxide*** which accounts for about 55 % of the change in the intensity of the Earth's greenhouse effect. The contributions of the other gases are 25 % for ***chlorofluorocarbons***, 15 % for ***methane***, and 5 % for ***nitrous oxide***. ***Ozone's*** contribution to the **enhancement** of greenhouse effect is still yet to be quantified.

Concentrations of ***carbon dioxide*** in the atmosphere are now approaching 360 parts per million (see ***Figure 7a-1***). Prior to 1700, levels of carbon dioxide were about 280 parts per million. This increase in carbon dioxide in the atmosphere is primarily due to the activities of humans. Beginning in 1700, societal changes brought about by the ***Industrial Revolution*** increased the amount of carbon dioxide entering the atmosphere. The major sources of this gas include fossil fuel combustion for industry, transportation, space heating, electricity generation and cooking; and vegetation changes in natural prairie, woodland, and forested ecosystems. Emissions from fossil fuel combustion account for about 65 % of the extra carbon dioxide now found in our atmosphere. The remaining 35 % is derived from deforestation and the conversion of prairie, woodland, and forested ecosystems primarily into agricultural systems. Natural ecosystems can hold 20 to 100 times more carbon dioxide per unit area than agricultural systems.

Artificially created ***chlorofluorocarbons*** are the strongest greenhouse gas per molecule. However, low concentrations in the atmosphere reduce their overall importance in the **enhancement** of the greenhouse effect. Current measurements in the atmosphere indicate that the concentration of these chemicals may soon begin declining because of reduced emissions. Reports of the development of ozone holes over the North and South Poles and a general decline in global stratospheric ozone levels over the last two decades has caused many nations to cutback on their production and use of these chemicals. In 1987, the signing of the ***Montreal Protocol*** agreement by forty-six nations established an immediate timetable for the global reduction of chlorofluorocarbons production and use.

Since 1750, ***methane*** concentrations in the atmosphere have increased by more than 140 %. The primary sources for the additional methane added to the atmosphere (in order of importance) are rice cultivation, domestic grazing animals, termites, landfills, coal mining, and oil and gas extraction. Anaerobic conditions associated with rice paddy flooding results in the formation of methane gas. However, an accurate estimate of how much methane is being produced from rice paddies has been difficult to obtain. More than 60 % of all rice paddies are found in India and China where scientific data concerning emission rates are unavailable. Nevertheless, scientists believe that the contribution of rice paddies is large because this form of crop production has more than doubled since 1950. Grazing animals release methane to the environment as a result of herbaceous digestion. Some researchers believe the addition of methane from this source has more than quadrupled over the last century. Termites also release methane through similar processes. Land-use change in the tropics, due to deforestation, ranching, and farming, may be causing termite numbers to expand. If this assumption is correct, the contribution from these insects may be important. Methane is also released from landfills, coal mines, and gas and oil drilling. Landfills produce methane as organic wastes decompose over time. Coal, oil, and natural gas deposits release methane to the atmosphere when these deposits are excavated or drilled.

The average concentration of ***nitrous oxide*** in the atmosphere is now increasing at a rate of 0.2 to 0.3 % per year. Sources for this increase include land-use conversion; fossil fuel combustion; biomass burning; and soil fertilization. Most of the nitrous oxide added to the atmosphere each year comes from deforestation and the conversion of forest, savanna and grassland ecosystems into agricultural fields and rangeland. Both of these processes reduce the amount of nitrogen stored in living vegetation and soil through the decomposition of organic matter. Nitrous oxide is also released into the atmosphere when fossil fuels and biomass are burned. However, the combined contribution of these sources to the increase of this gas in the atmosphere is thought to be minor. The use of nitrate and ammonium fertilizers to enhance plant growth is another source of nitrous oxide. Accurate measurements of how much nitrous oxide is being released from fertilization have been difficult to obtain. Estimates suggest that the contribution from this source may represent from 50 % to 0.2 % of nitrous oxide added to the atmosphere annually.

Ozone's role in the enhancement of the greenhouse effect has been difficult to determine scientifically. Accurate measurements of past long-term (more than 25 years in the past) levels of this gas in the atmosphere are currently unavailable. Concentrations of ozone gas are found in two different regions of the Earth's atmosphere. The majority of the ozone (about 97 %) found in the atmosphere is localized in the ***stratosphere*** at an altitude of 15 to 55 kilometers above the Earth's surface. In recent years, the concentration of the ***stratospheric ozone*** has been decreasing because of the buildup of ***chlorofluorocarbons*** in the atmosphere (see ***Lecture 7e***). Since the late 1970s, scientists have discovered that ***total column ozone*** amounts over Antarctica in the springtime have decreased by as much as 70 %. Satellite measurements have indicated that the zone from 65 degrees North to 65 degrees South latitude has had a 3 % decrease in stratospheric ozone since 1978. Ozone is also highly concentrated at the Earth's surface. Most of this ozone is created as an artificial by product of ***photochemical smog***.

In summary, the **greenhouse effect** causes the atmosphere to trap more heat energy at the Earth's surface and within the atmosphere by absorbing and re-emitting longwave energy. Of the longwave energy emitted back to space, 90 % is intercepted and absorbed by greenhouse gases. Without the greenhouse effect the Earth's average global temperature would be -18 degrees Celsius, rather than the present 15 degrees Celsius. In the last few centuries, the activities of humans have directly or indirectly caused the concentration of the major greenhouse gases to increase. Scientists predict that this increase may **enhance** the greenhouse effect making the planet warmer. Some experts estimate that the Earth's average global temperature has already increased by 0.3 to 0.6 degrees Celsius, since the beginning of this century, because of this enhancement. Predictions of future climates indicate that by the middle of the next century the Earth's global temperature may be 1 to 3 degrees Celsius higher than today.

Table 7h-1: Gases involved in the Greenhouse Effect: past and present concentration and sources.

{PRIVATE}**Greenhouse Gas**	**Concentration 1750**	**Present (1995) Concentration**	**Percent Change**	**Natural and Anthropogenic Sources**
Carbon Dioxide	280 ppm	360 ppm	29 %	Organic decay; Forest fires; Volcanoes; Burning fossil fuels; Deforestation; Land-use change
Methane	0.70 ppm	1.70 ppm	143 %	Wetlands; Organic decay; Termites; Natural gas & oil extraction; Biomass burning; Rice cultivation; Cattle; Refuse landfills
Nitrous Oxide	280 ppb	310 ppb	11 %	Forests; Grasslands; Oceans; Soils; Soil cultivation; Fertilizers; Biomass burning; Burning of fossil fuels
Chlorofluorocarbons (CFCs)	0	900 ppt	Not Applicable	Refrigerators; Aerosol spray propellants; Cleaning solvents
Ozone	Unknown	Varies with latitude and altitude in the atmosphere	Global levels have generally decreased in the stratosphere and increased near the Earth's surface	Created naturally by the action of sunlight on molecular oxygen and artificially through photochemical smog production

(i) Net Radiation and the Planetary Energy Balance

{PRIVATE}***Shortwave radiation*** from the sun enters the surface-atmosphere system of the Earth and is ultimately returned to space as ***longwave radiation*** (because the Earth is cooler than the sun). A basic necessity of this energy interchange is that incoming solar insolation and outgoing radiation be equal in quantity. One way of modeling this balance in energy exchange is described graphically with the use of the following two **cascade** diagrams.

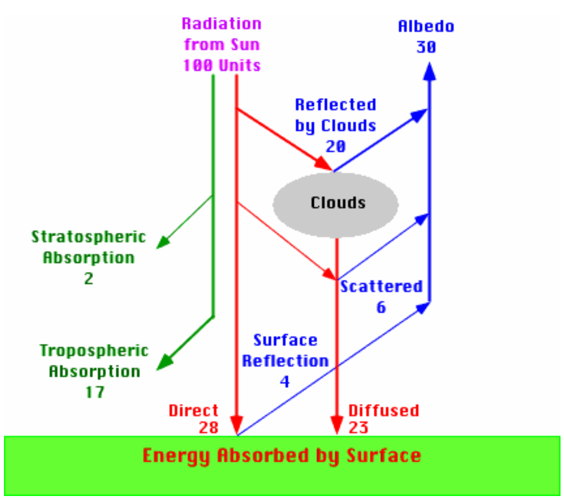


Figure 7i-1: Global shortwave radiation cascade.

The **Global Shortwave Radiation Cascade** describes the relative amounts (based on **100 units** available at the top of the atmosphere) of ***shortwave radiation*** partitioned to various atmospheric processes as it passes through the atmosphere. The diagram indicates that **19 units** of insolation are ***absorbed*** (and therefore transferred into ***heat energy*** and ***longwave radiation***) in the atmosphere by the following two processes:

Stratospheric Absorption of the Ultraviolet Radiation by Ozone 2 units; and

Tropospheric Absorption of Insolation by Clouds and Aerosols 17 units.

23 units of solar radiation are ***scattered*** in the atmosphere subsequently absorbed at the surface as ***diffused insolation***. **28 units** of the incoming solar radiation are absorbed at the surface as ***direct insolation***. Total amount of solar insolation absorbed at the surface equals **51 units**. The total amount of shortwave radiation absorbed at the surface and in the atmosphere is **70 units**.

Three main losses of solar radiation back to space occur in the Earth's shortwave radiation cascade. **4 units** of sunlight are returned to space from ***surface reflection***. Cloud reflection returns another **20 units** of solar radiation. Back ***scattering*** of sunlight returns **6 units** to space. The total loss of shortwave radiation from these processes is **30 units**. The term used to describe the combined effect of all of these shortwave losses is ***Earth albedo***.

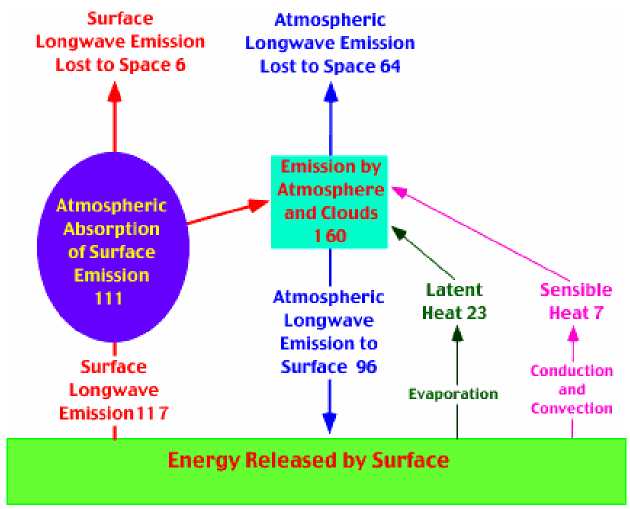


Figure 7i-2: Global longwave radiation cascade.

The **Global Longwave Radiation Cascade** indicates that energy leaves the Earth's surface through three different processes. **7 units** leave the surface as ***sensible heat***. This heat is transferred into the atmosphere by ***conduction*** and ***convection***. The melting and evaporation of water at the Earth's surface incorporates **23 units** energy into the atmosphere as ***latent heat***. This ***latent heat*** is released into the atmosphere when the water condenses or becomes solid. Both of these processes become part of the emission of longwave radiation by the atmosphere and clouds.

The surface of the Earth emits **117 units** of ***longwave radiation***. Of this emission only **6 units** are directly lost to space. The other **111 units** are absorbed by ***greenhouse gases*** in the atmosphere and converted into ***heat energy*** and then into atmospheric emissions of longwave radiation (the ***greenhouse effect***).

The atmosphere emits **160 units** of longwave energy. Contributions to this **160 units** are from surface emissions of ***longwave radiation*** (**111 units**), ***latent heat*** transfer (**23 units**), ***sensible heat*** transfer (**7 units**), and the absorption of ***shortwave radiation*** by atmospheric gases and clouds (**19 units**, see **Figure 7i-1**). Atmospheric emissions travel in two directions. **64 units** of atmospheric emission is lost directly to space. **96 units** travel to the Earth's surface where it is absorbed and transferred into ***heat energy***.

The total amount of energy lost to space in the global longwave radiation cascade is **70 units** (surface emission **6 units** + atmospheric emission **64 units**.) This is the same amount of energy that was added to the Earth's atmosphere and surface by the **Global Shortwave Radiation Cascade**.

Finally, to balance the surface energy exchanges in this cascade we have to account for **51 units** of missing energy [atmosphere and cloud longwave emission (**96 units**) **minus** surface longwave emission (**117 units**) **minus** ***latent heat*** transfer (**23 units**) **minus** ***sensible heat*** transfer (**7 units**) = **-51 units**]. This missing component to the radiation balance is the **51 units** of energy absorbed at the Earth's surface as ***direct*** and ***diffused shortwave radiation*** (see **Figure 7i-1**).

The following equations can be used to mathematically model **net shortwave radiation balance**, **net longwave radiation balance**, and **net radiation balance** for the Earth's surface at a single location or for the whole globe for any temporal period:

$$\mathbf{K^*} = (\mathbf{K} + \mathbf{k})(1 - \mathbf{a})$$

$$\mathbf{L^*} = (\mathbf{LD} - \mathbf{LU})$$

$$\mathbf{Q^*} = (\mathbf{K} + \mathbf{k})(1 - \mathbf{a}) - \mathbf{LU} + \mathbf{LD}$$

where

Q* is surface **net radiation** (global annual values of **Q* = 0** , because input equals output, local values can be positive or negative),

K*is surface **net shortwave radiation**,

K is surface ***direct shortwave radiation***,

k is ***diffused shortwave radiation*** (***scattered*** insolation) at the surface,

a is the ***albedo*** of surface,

L*is **net longwave radiation** at the surface,

LD is atmospheric ***counter-radiation*** (see ***Greenhouse Effect***) directed to the Earth's surface, and

LU is ***longwave radiation*** lost from the Earth's surface.

NASA's *Surface Radiation Budget Project* has used satellite data, computer models, and meteorological data to determine surface net shortwave radiation, net longwave radiation, and net radiation balances for the period July 1983 to June 1991. The following **links** display these balances for January and July globally:

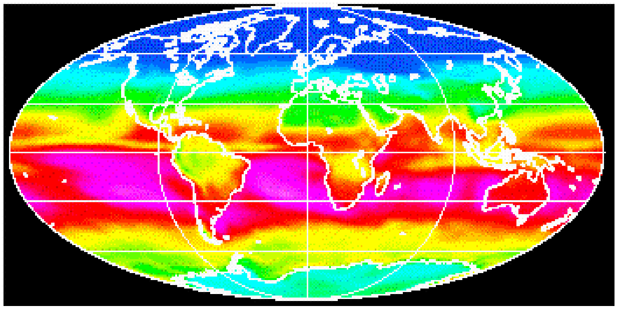


Figure 7i-3: Average net shortwave radiation at the Earth's surface: January 1984-1991. Highest values occur along the subtropical oceans of the Southern Hemisphere. Lowest values occur over areas of high surface reflection such as the South Pole, cloudy regions, and areas of low solar input like the high latitudes of the Northern Hemisphere. Color range: blue - red - white, Values: 0 to 350W/m**2. Global mean = 162W/m**2, Minimum = 0W/m**2, Maximum = 315W/m**2. (**Source:** NASA *Surface Radiation Budget Project*).***Average Net Shortwave Radiation at the Earth's Surface: January 1984-1991*** **(K*)**

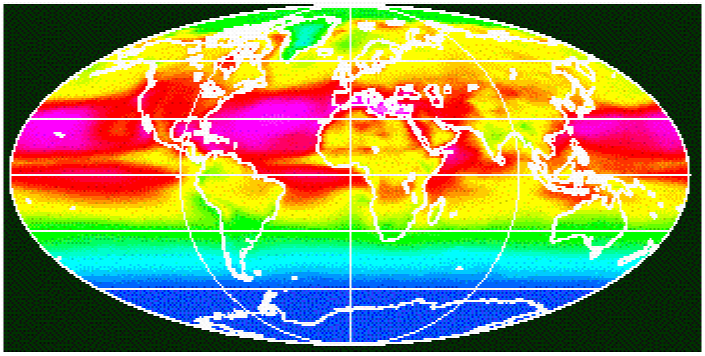


Figure 7i -4: Average net shortwave radiation at the Earth's surface: July 1983-1990. Highest values occur over the subtropical oceans of the Northern Hemisphere due to high solar input and little cloud coverage. Lowest values occur in areas of high surface reflection such as snow/ice covered surfaces like Greenland, cloudy regions such as storm tracks, and areas of low solar input like the high latitudes of the Southern Hemisphere. Color range: blue - red - white, Values: 0 to 350W/m**2. Global mean = 158/m**2, Minimum = 0W/m**2, Maximum = 323/m**2. (**Source:** NASA *Surface Radiation Budget Project*).***Average Net Shortwave Radiation at the Earth's Surface: July 1983-1990*** **(K*)**

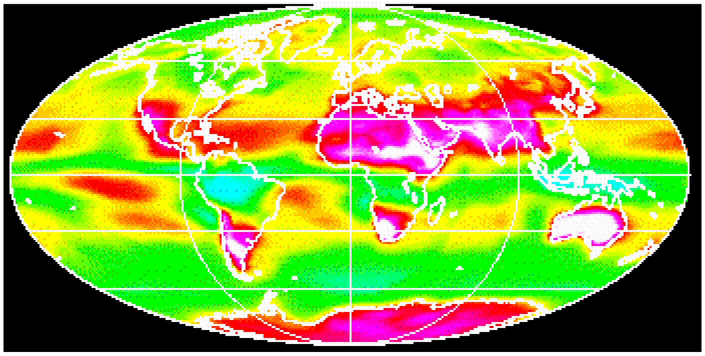


Figure 7i -5: Average net longwave radiation at the Earth's surface: January 1984-1991. Net longwave loss is a negative quantity. Highest values of longwave loss occurs where surface temperatures are high and cloud cover is minimal, such as the subtropical deserts of the Northern and Southern Hemisphere. Cold surfaces have low values of loss. Color range: white - red - blue, Values: -100 to 0W/m**2. Global mean = -48W/m**2, Minimum = -125W/m**2, Maximum = -11W/m**2. (**Source:** NASA *Surface Radiation Budget Project*). ***Average Net Longwave Radiation at the Earth's Surface: January 1984-1991*** **(L*)**

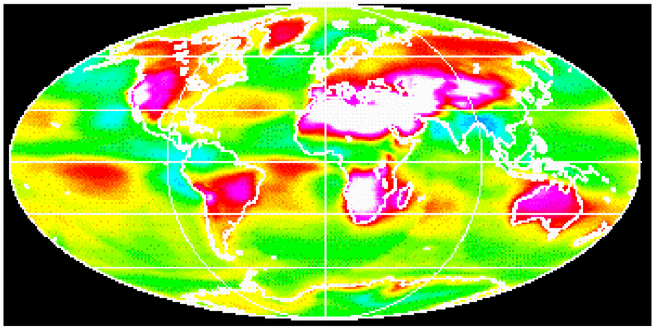


Figure 7i -6: Average net longwave radiation at the Earth's surface: July 1983-1990. Net longwave loss is a negative quantity. Highest values of longwave loss occurs where surface temperatures are high and cloud cover is minimal, such as the subtropical deserts of the Northern and Southern Hemisphere. Cold surfaces have low values of loss. Color range: white - red - blue, Values: -100 to 0W/m**2. Global mean = -47W/m**2, Minimum = -144W/m**2, Maximum = -4W/m**2. (**Source:** NASA *Surface Radiation Budget Project*). ***Average Net Longwave Radiation at the Earth's Surface: July 1983-1990*** **(L*)**

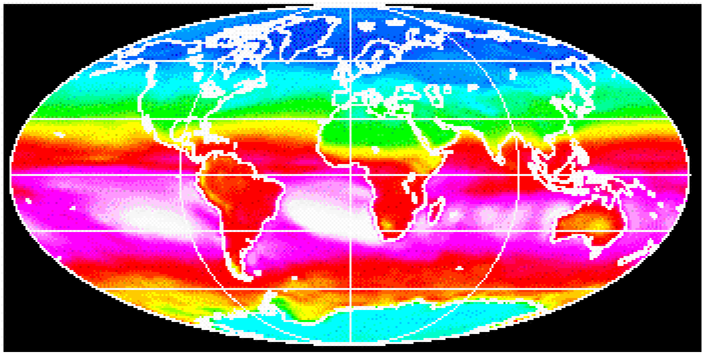


Figure 7i -7: Average net radiation at the Earth's surface: January 1984-1991. Total net radiation is the sum of shortwave and longwave net radiation. It is dominated by the shortwave portion. Highest values occur along the subtropical oceans of the Southern Hemisphere. Lowest values occur over areas of low solar input such as the North Pole, and areas of high surface reflection such as the South Pole. Color range: blue - red - white, light green = 0 W/m**2, Values: -50 to 250W/m**2. Global mean = 114W/m**2, Minimum = -60W/m**2, Maximum = 261W/m**2. (**Source:** NASA *Surface Radiation Budget Project*). ***Average Net Radiation at the Earth's Surface: January 1984-1991*** **(Q*)**

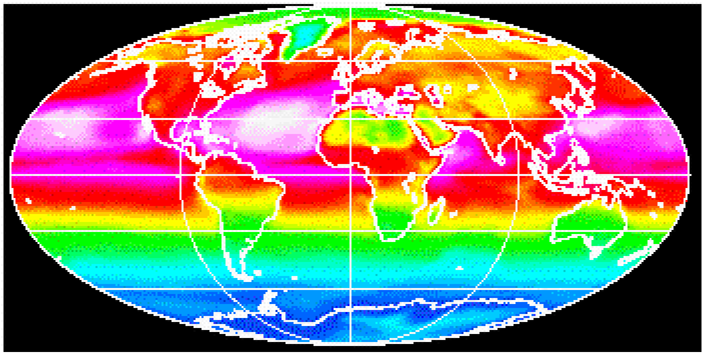
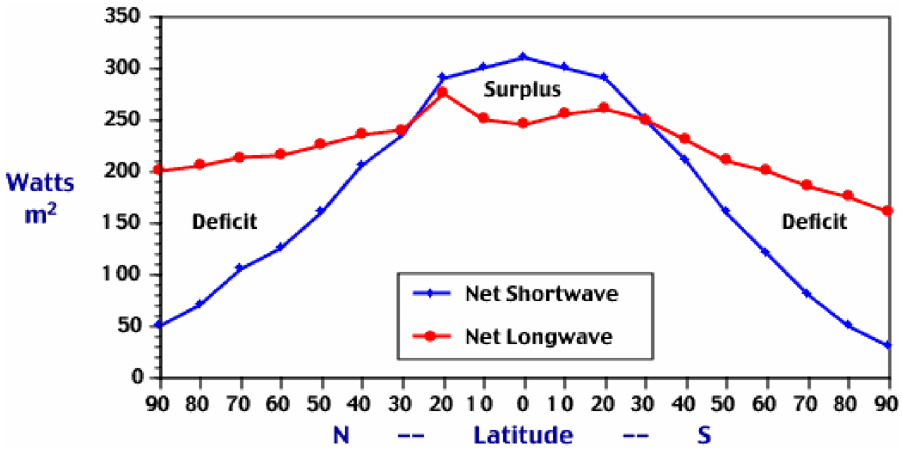


Figure 7i -8: Average net radiation at the Earth's surface: July 1983-1990. Total net radiation is the sum of shortwave and longwave net radiation. It is dominated by the shortwave portion. Highest values occur along the subtropical oceans of the Northern Hemisphere. Lowest values occur over areas of low solar input such as the South Pole, and areas of high surface reflection such as the North Pole. Color range: blue - red - white, light green = 0 W/m**2, Values: -50 to 250W/m**2. Global mean = 111W/m**2, Minimum = -65W/m**2, Maximum = 249W/m**2. (**Source:** NASA *Surface Radiation Budget Project*). ***Average Net Radiation at the Earth's Surface: July 1983-1990*** **(Q*)**

(j) Global Heat Balance: Introduction to Heat Fluxes

{PRIVATE**}Figure 7j-1** illustrates the annual values of net shortwave and net longwave radiation from the South Pole to the North Pole. On closer examination of this graph one notes that the lines representing **incoming** and **outgoing radiation** do not have the same values. From 0 - 30 degrees latitude North and South incoming solar radiation exceeds outgoing terrestrial radiation and a surplus of energy exists. The reverse holds true from 30 - 90 degrees latitude North and South and these regions have a deficit of energy. Surplus energy at low latitudes and a deficit at high latitudes results in energy transfer from the equator to the poles. It is this ***meridional transport*** of energy that causes atmospheric and oceanic circulation. If there were no energy transfer the poles would be 25 degrees Celsius cooler, and the equator 14 degrees Celsius warmer!



{PRIVATE}**Figure 7j-1:** Balance between average net shortwave and longwave radiation from 90 degrees North to 90 degrees South.

The redistribution of energy across the Earth's surface is accomplished primarily through three processes: ***sensible heat flux***, ***latent heat flux***, and ***surface heat flux*** into oceans. ***Sensible heat flux*** is the process where heat energy is transferred from the Earth's surface to the atmosphere by ***conduction*** and ***convection***. This energy is then moved from the tropics to the poles by ***advection***, creating atmospheric circulation. As a result, atmospheric circulation moves warm tropical air to the polar regions and cold air from the poles to the equator. ***Latent heat flux*** moves energy globally when solid and liquid water is converted into vapor. This vapor is often moved by atmospheric circulation vertically and horizontally to cooler locations where it is ***condensed*** as rain or is ***deposited*** as snow releasing the heat energy stored within it. Finally, large quantities of radiation energy are transferred into the Earth's tropical oceans. The energy enters these water bodies at the surface when absorbed radiation is converted into heat energy. The warmed surface water is then transferred downward into the water column by conduction and convection. Horizontal transfer of this heat energy from the equator to the poles is accomplished by ***ocean currents***.

The following equation describes the partitioning of heat energy at the Earth's surface:

Q* = H (***Sensible heat***) **+ L** (***Latent heat***) **+ S** (***Surface heat flux*** into soil or water)

The actual amount of net radiation being partitioned into each one of these components is a function of the following factors:

Presence or absence of water in liquid and solid forms at the surface.

Specific heat of the surface receiving the ***net radiation***.

Convective and ***conductive*** characteristics of the receiving surface.

Diffusion characteristics of the surface's overlying atmosphere.

(k) The Concept of Temperature

{PRIVATE}Temperature and Heat

Temperature and ***heat*** are not the same phenomenon. ***Temperature*** is a measure of the intensity or degree of hotness in a body. Technically, it is determined by getting the average speed of a body's molecules. ***Heat*** is a measure of the quantity of heat energy present in a body. The spatial distribution of temperature in a body determines heat flow. Heat always flows from warmer to colder areas.

The heat held in a object depends not only on its ***temperature*** but also its ***mass***. For example, let us compare the heating of two different masses of water (**Table 7k-1**). In this example, one mass has a weight of 5 grams, while the other is 25 grams. If the temperature of both masses is raised from 20 to 25 degrees Celsius, the larger mass of water will require five times more heat energy for this increase in temperature. This larger mass would also contain contain 5 times more stored heat energy.

{PRIVATE**}Table 7k-1**: Heat energy required to raise two different quantities of water 5 degrees Celsius.

{PRIVATE**}Mass of the Water**	**Starting Temperature**	**Ending Temperature**	**Heat Required**
5 grams	20 degrees Celsius	25 degrees Celsius	25 Calories of Heat
25 grams	20 degrees Celsius	25 degrees Celsius	125 Calories of Heat

Temperature Scales

A number of measurement scales have been invented to measure temperature. **Table 7k-2** describes important temperatures for the three dominant scales in use today.

{PRIVATE**}Table 7k-2**: Temperature of absolute zero, the ice point of water, and the stream point of water using various temperature measurement scales.

{PRIVATE**}Measurement Scale**	**Steam Point of Water**	**Ice Point of Water**	**Absolute Zero**
Fahrenheit	**212**	**32**	**-460**
Celsius	**100**	**0**	**-273**
Kelvin	**373**	**273**	**0**

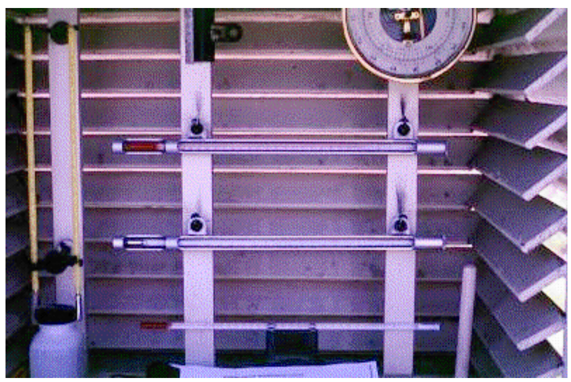
The most commonly used scale for measuring temperature is the ***Celsius*** system. The Celsius scale was developed in 1742 by the Swedish astronomer Anders Celsius. In this system, the melting point of ice was given a value of 0, the boiling point of water is 100, and ***absolute zero*** is -273. The ***Fahrenheit*** system is a temperature scale that is used exclusively in the United States. This system was created by German physicist Gabriel Fahrenheit in 1714. In this scale, the melting point of ice has a value of 32, water boils at 212, and absolute zero has a temperature of -460. The ***Kelvin*** scale was proposed by British physicist Lord Kelvin in 1848. This system is often used by scientists because its temperature readings begin at absolute zero and due to the fact that this scale is proportional to the amount of heat energy found in an object. The Kelvin scale assigns a value of 273 for the melting temperature of ice, while the boiling point of water occurs at 373.

Measurement of Air Temperature

A ***thermometer*** is a device that is used to measure temperature. Thermometers consist of a sealed hollow glass tube filled with some type of liquid. Thermometers measure temperature by the change in the volume of the liquid as it responds to the addition or loss of heat energy from the environment immediately outside its surface. When heat is added, the liquid inside the thermometer expands. Cooling cause the liquid to contract. Meteorological thermometers are often filled with either alcohol or mercury. Alcohol thermometers are favored in very cold environments because of this liquid's low freezing point (-112 degrees Celsius).

By international agreement, the nations of the world have decided to measure temperature in a similar fashion. This standardization is important for the accurate generation of weather maps and forecasts, both of which depend on having data determined in a uniform way. Weather stations worldwide try to determine minimum and maximum temperatures for each and every day. By averaging these two values, daily mean temperatures are also calculated. Many stations also take temperature readings on the hour. Temperature measurements are determined by thermometers designed and approved by the World Meteorological Organization (see *http://www.wmo.ch*). These instruments are housed in specially designed instrument shelters that allow for the standardization of measurements taken anywhere on the Earth (**Figure 7k-1** and **Figure 7k-2**).

{PRIVATE**}Figure 7k-1:** Well ventilated instrument shelters are used to protect thermometers from precipitation, direct sun, and other physical elements. Construction standardization of these shelters, by international agreement, guarantees that measurements are comparable in any of the over 15,000 weather stations found worldwide.



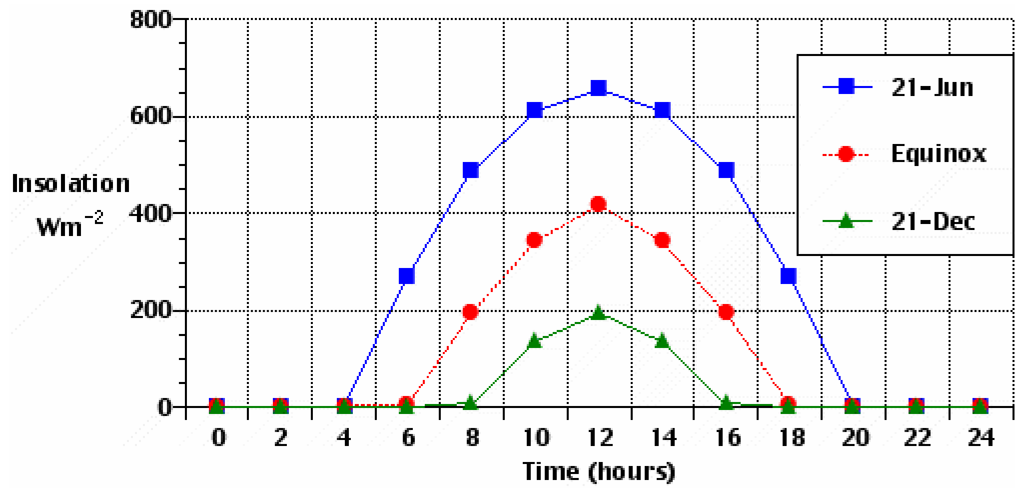
{PRIVATE**}Figure 7k-2:** Thermometers found inside the instrument shelter are mounted approximate 1.5 meters above the ground surface. The top thermometer contains alcohol and is used to determine daily minimum temperatures. The lower thermometer uses mercury to determine the daily maximum temperature.

(l) Daily and Annual Cycles of Temperature

{PRIVATE**}Daily Cycles of Air Temperature**

At the Earth's surface quantities of ***insolation*** and ***net radiation*** undergo daily cycles of change because the planet rotates on its polar axis once every 24 hours. Insolation is usually the main positive component making up net radiation. Variations in net radiation are primarily responsible for the particular patterns of rising and falling air temperature over a 24 hour period. The following three **graphs** show hypothetical average curves of **insolation**, **net radiation**, and **air temperature** for a typical land based location at 45 degrees of latitude on the ***equinoxes*** and ***solstices*** (**Figures 7l-1**, **7l-2**, and **7l-3**).

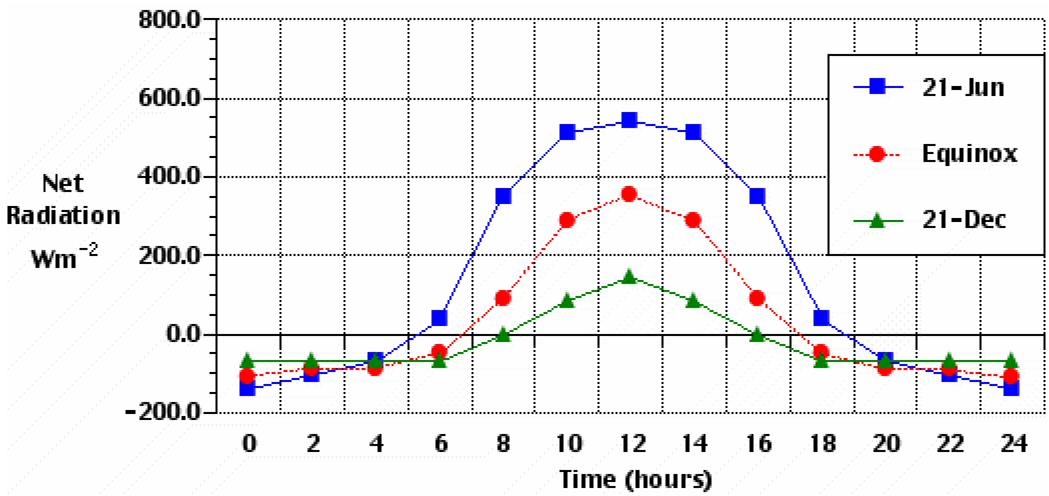
Insolation



{PRIVATE}Figure 7l-1: Hourly variations in insolation received for a location at 45 degrees North latitude over a 24 hour period.

In the above graph, ***shortwave radiation*** received from the sun is measured in ***Watts***. For all dates, peak reception occurs at solar noon when the sun attains its greatest height above the horizon.

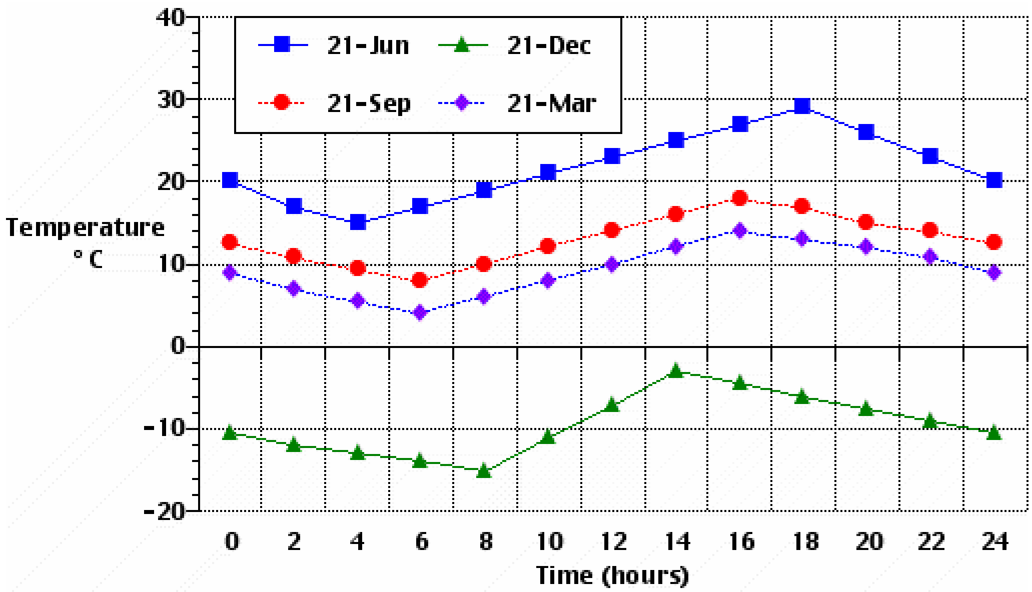
Net Radiation



{PRIVATE}Figure 7l-2: Hourly variations in net radiation for a location at 45 degrees North latitude over a 24 hour period.

Units in **Figure 7l-2** are the same as the ***insolation*** graph above. The ***net radiation*** graph indicates that there is a surplus of radiation during most of the day and a deficit throughout the night. The deficit begins just before sunset when emitted longwave radiation from the Earth's surface exceeds solar insolation and longwave radiation from the atmosphere.

Temperature



{PRIVATE**}Figure 7l-3:** Hourly variations in surface temperature for a location at 45 degrees North latitude over a 24 hour period.

The relative placement of the temperature profiles for the various dates correlates to the amount of ***net radiation*** available for daily surface absorption and heat generation. The more energy available, the higher up the ***Y-axis*** the profile is on the **graph**. ***Autumnal equinox*** (September 21) is warmer than the ***vernal equinox*** (March 21) because of the heating that occurred in the previous summer months. For all dates, minimum temperature occurs at **sunrise**. Temperature drops throughout the night because of two processes. First, the Earth's radiation balance at the surface becomes negative after **sunset**. Thus, the surface of the Earth stops heating up as solar radiation is not being absorbed. Secondly, ***conduction*** and ***convection*** transport heat energy up into the atmosphere and the warm air that was at the surface is replaced by cooler air from above because of atmospheric mixing. Temperature begins rising as soon as the net radiation budget of the surface becomes positive. Temperature continues to rise from sunrise until sometime after solar noon. After this time, mixing of the Earth's surface by convection causes the surface to cool despite the positive addition of radiation and ***heat energy***.

Annual Cycle of Air Temperature

As the Earth revolves around the sun, locations on the surface may under go seasonal changes in air temperature because of annual variations in the intensity of ***net radiation***. Variations in net radiation are primarily controlled by changes in the intensity and duration of received solar insolation which are driven by variations in daylength and angle of incidence. The discussion below examines how changes in net radiation can effect **mean monthly temperatures** for the following five locations:

Manaus, Brazil, 3 degrees South latitude (**Figure 7l-4**).

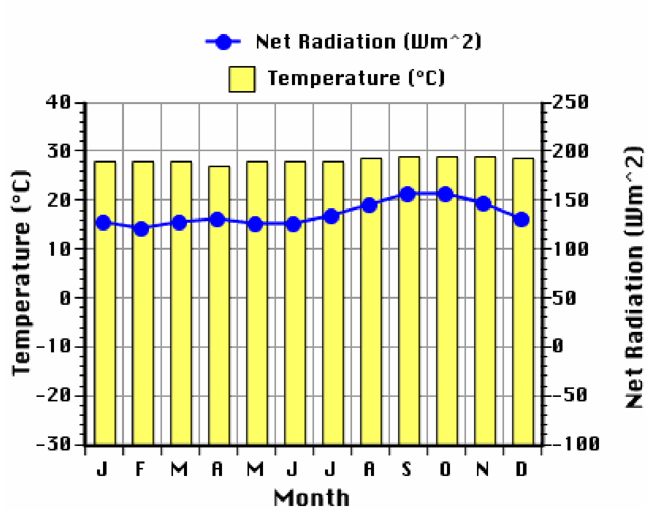
Bulawayo, Zimbabwe, 20 degrees South latitude (**Figure 7l-5**).

Albuquerque, USA, 35 degrees North latitude (**Figure 7l-6**).

London, England, 52 degrees North latitude (**Figure 7l-7**).

Fairbanks, USA, 65 degrees North latitude (**Figure 7l-8**).

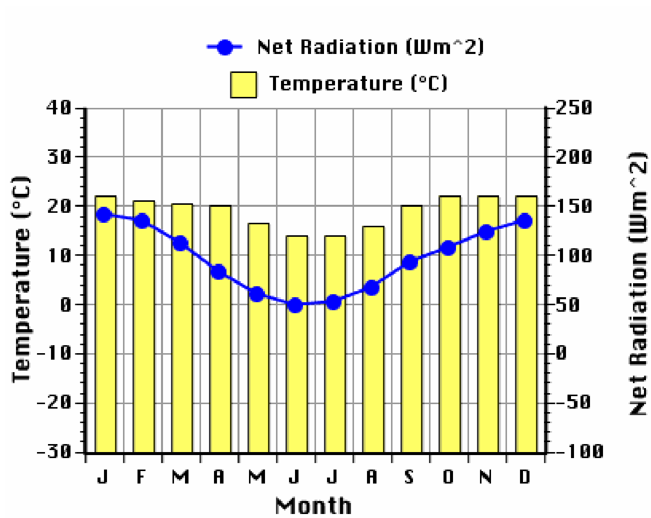
Manaus, Brazil - 3°South, 60°West



{PRIVATE**}Figure 7l-4:** Monthly variations in net radiation and average monthly temperature for Manaus, Brazil.

At **Manaus**, values of monthly net radiation average about 135 ***Watts*** per square meter. Monthly variation in net radiation is only about 35 ***Watts*** over the entire year (**Figure 7l-4**). Two peaks in net radiation are visible on the graph. Both of these peaks occur during the ***equinoxes*** when the height of the sun above the ***horizon*** is at its maximum (90 degrees above the horizon). Minimum values of net radiation correspond to the time of the year when the sun reaches its minimum height of only 66.5 degrees above the horizon at ***solar noon***. Because of the consistent nature of net radiation, mean monthly air temperature only varies by 2 degrees Celsius over the entire year.

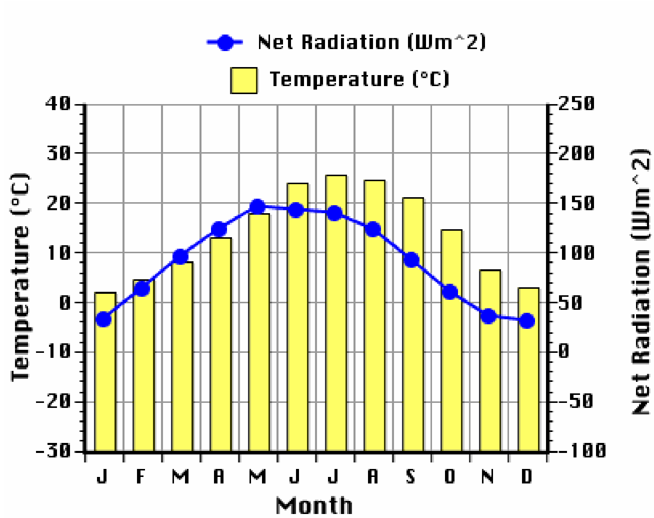
Bulawayo, Zimbabwe - 20°South, 29°East



{PRIVATE**}Figure 7l-5:** Monthly variations in net radiation and average monthly temperature for Bulawayo, Zimbabwe.

Net radiation at **Bulawayo** has a single peak and trough over the one year period graphed. This pattern is primarily controlled by variations in the intensity and duration of incoming solar insolation (**Figure 7l-5**). During the ***winter solstice*** the sun reaches its highest altitude above the horizon and daylength is at a maximum (13 hours and 12 minutes). The lowest values of net radiation occur around the ***summer solstice*** when the sun reaches its lowest altitude above the horizon and daylength is at a minimum (10 hours and 48 minutes) in the Southern Hemisphere. Monthly temperature variations follow the monthly change in net radiation. Net radiation represents energy available to do work. When received at the Earth's surface much of this energy is used to create ***sensible heat***.

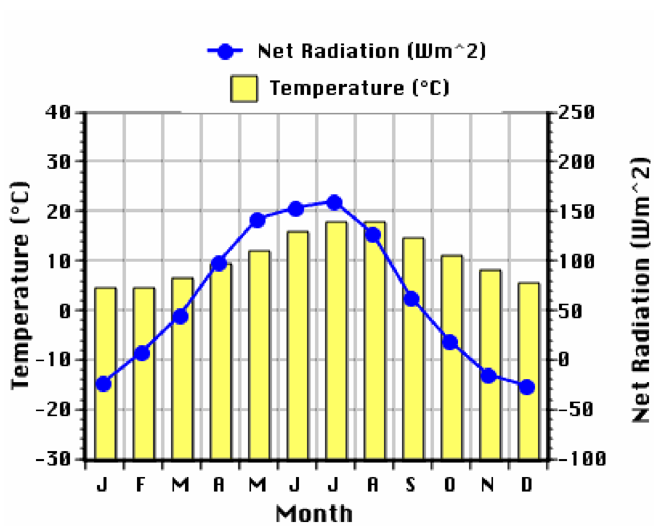
Albuquerque, USA - 35°North, 107°West



{PRIVATE**}Figure 7l-6:** Monthly variations in net radiation and average monthly temperature for Albuquerque, USA.

At **Albuquerque**, maximum net radiation occurs in May. The timing of this peak roughly coincides with the ***summer solstice*** when daylengths are at their longest and solar heights are their greatest (**Figure 7l-6**). However, monthly temperature variations do not mirror the changes in net radiation exactly. Peak monthly temperatures occur about two months after the net radiation maximum. This lag is probably caused by the delayed movement of stored heat energy in the ground into the atmosphere. Minimum monthly temperatures do coincide with the lowest values of net radiation which occur during the ***winter solstice***.

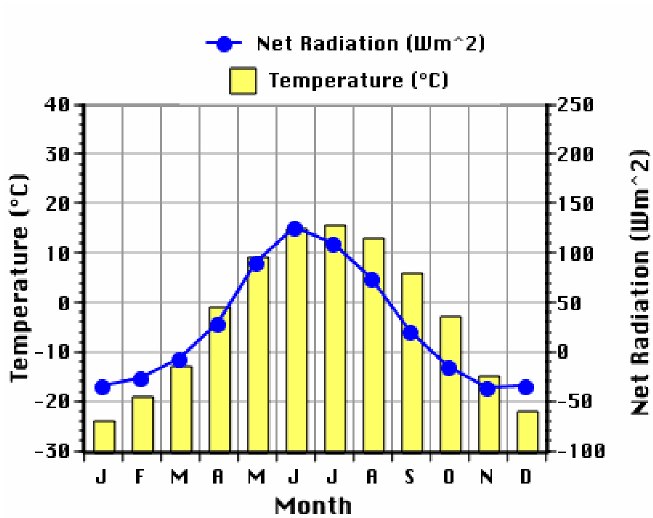
London, England - 52°North, 1°East



{PRIVATE**}Figure 7l-7:** Monthly variations in net radiation and average monthly temperature for London, England.

The annual patterns of net radiation and mean monthly temperature for **London** are quite similar to those already described for Albuquerque (**Figures 7l-6** and **7l-7**). London does, however, experience a greater annual variation in net radiation. This greater variation can be explained by the effect increasing latitude has on annual variations of insolation. During the winter months, outgoing ***longwave radiation*** actually exceeds incoming ***insolation*** producing negative net radiation values. This was not seen in Albuquerque. The variation in monthly mean temperature is also less extreme in London when compared to Albuquerque. Intuitively, one would expect London to have a greater annual change in temperature because of the greater variation in net radiation over the year. However, London's climate is moderated by the frequent addition of ***latent heat*** energy from seasonal precipitation.

Fairbanks, USA - 65°North, 148°West



{PRIVATE**}Figure 7l-8:** Monthly variations in net radiation and average monthly temperature for Fairbanks, USA.

Of the five locations examined, **Fairbanks** has the greatest variations in mean monthly temperature. Fairbanks is also the coldest of the climates examined (**Figure 7l-8**). This is primarily due to the fact that during six months of the year net radiation is negative because outgoing ***longwave radiation*** exceeds incoming ***insolation***. Fairbanks also receives the least cumulative amount of net radiation over the entire year. Mean month temperature is at its maximum in July which is one month ahead of the peak in net radiation.

(m) Global Surface Temperature Distribution

{PRIVATE**}**If the Earth was a homogeneous body without the present land/ocean distribution, its temperature distribution would be strictly latitudinal (**Figure 7m-1**). However, the Earth is more complex than this being composed of a mosaic of land and water. This mosaic causes latitudinal zonation of temperature to be disrupted spatially.

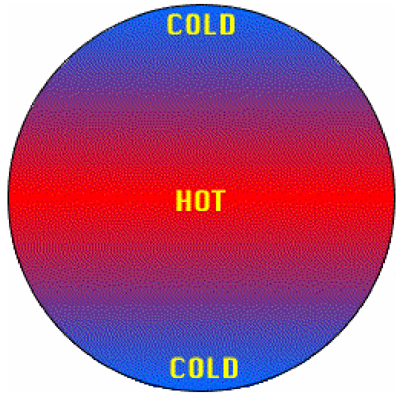


Figure 7m-1: Simple latitudinal zonation of temperature.

The following two factors are important in influencing the distribution of temperature on the Earth's surface:

The latitude of the location determines how much solar radiation is received. Latitude influences the angle of incidence and duration of daylength.

Surface properties - surfaces with high ***albedo*** absorb less incident radiation. In general, land absorbs less insolation that water because of its lighter color. Also, even if two surfaces have the same albedo, a surface's ***specific heat*** determines the amount of heat energy required for a specific rise in temperature per unit mass. The specific heat of water is some five times greater than that of rock and the land surface (see **Table 7m-1** below). As a result, water requires the input of large amounts of energy to cause a rise in its temperature.

Table 7m-1: Specific Heat of Various Substances.

{PRIVATE**}Substance**	**Specific Heat**
Water	1.00
Air	0.24
Granite	0.19
Sand	0.19
Iron	0.11

Mainly because of ***specific heat***, land surfaces behave quite differently from water surfaces. In general, the surface of any extensive deep body of water heats more slowly and cools more slowly than the surface of a large land body. Other factors influencing the way land and water surfaces heat and cool include:

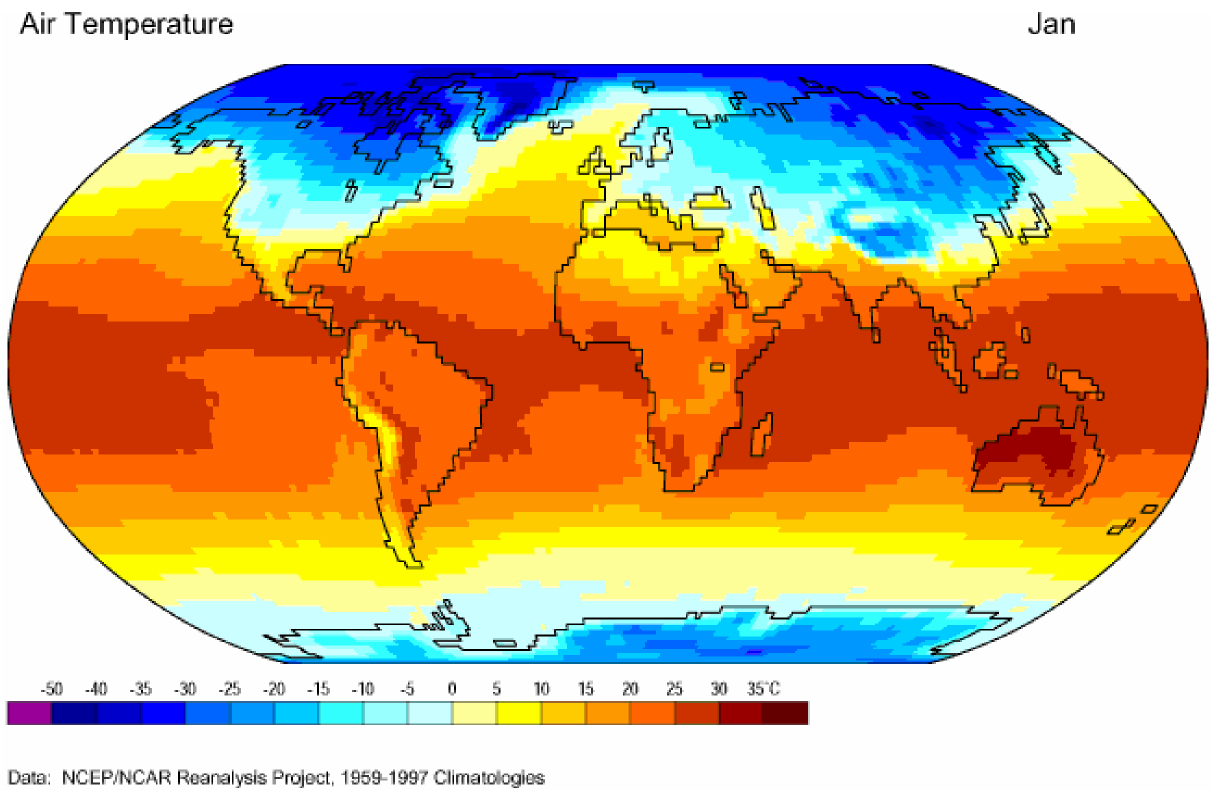
Solar radiation warms an extensive layer in water, on land just the immediate surface is heated.

Water is easily mixed by the process of ***convection***.

Evaporation of water removes energy from water's surface.

The following images illustrate the Earth's temperature distribution patterns for an average January and July based on 39 years of data (**Figures 7m-2** and **7m-3**). Note that the spatial variations of temperature on these figures is mostly latitudinal. However, the horizontal banding of ***isotherms*** is somewhat upset by the fact that water heats up more slowly in the summer and cools down more slowly in the winter when compared to land surfaces. During January, much of the terrestrial areas of the Northern Hemisphere are below freezing (**Figure 7m-2**). Some notable Northern Hemisphere cold-spots include the area around Baffin Island Canada, Greenland, Siberia, and the Plateau of Tibet. Temperatures over oceans tend to be hotter because of the water's ability to hold heat energy.

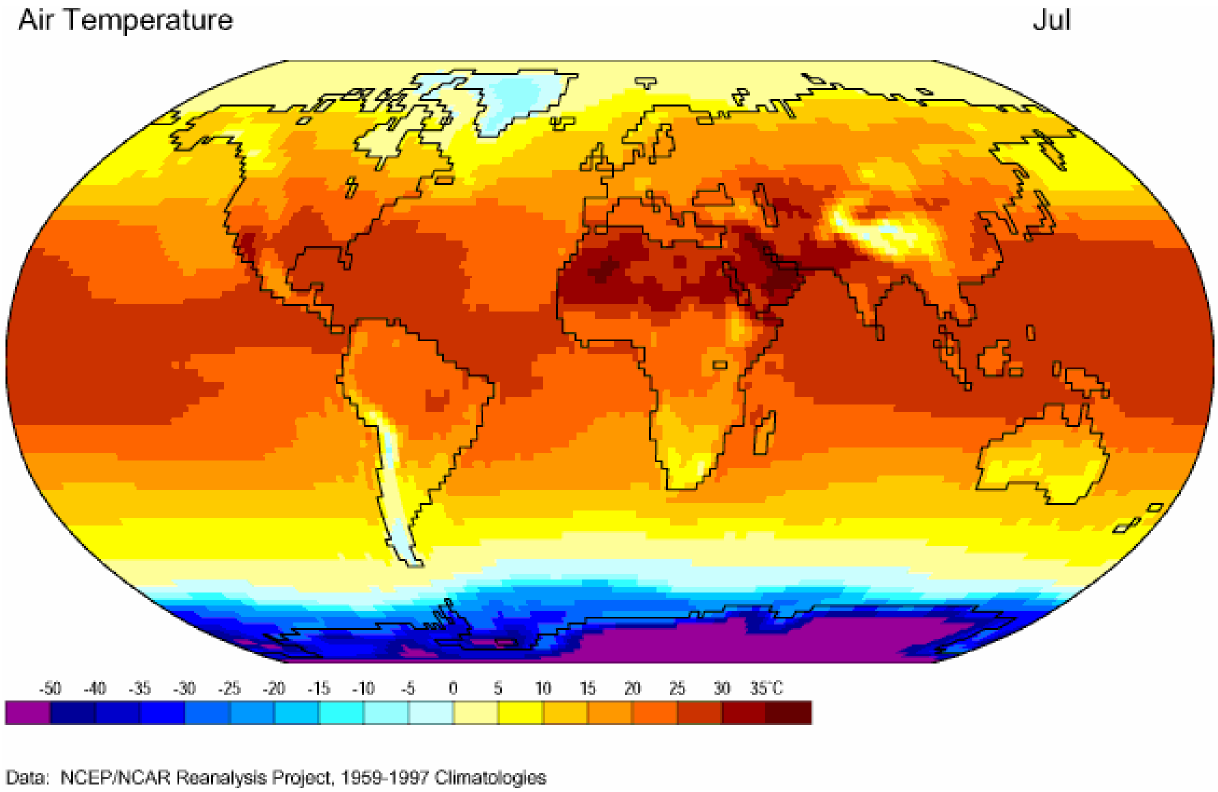
In the Southern Hemisphere, temperatures over the major landmasses are generally greater than 20 degrees Celsius with localized hot-spots in west-central Australia, the Kalahari Desert in Africa, and the plains of Bolivia, Paraguay, and Argentina (**Figure 7m-2**). Subtropical oceans are often warmer than landmass areas near the equator. At this latitude, land areas receive less incoming solar radiation because of the daily ***convective*** development of ***cumulus*** and ***cumulonimbus*** clouds. In the mid-latitudes, oceans are often cooler than landmass areas at similar latitudes. Terrestrial areas are warmer because of the rapid heating of land surfaces under frequently clear skies. Antarctica remains cold and below zero degrees Celsius due to the presence of permanent glacial ice which reflects much of the solar radiation received back to space.



{PRIVATE**}Figure 7m-2:** Mean January air temperature for the Earth's surface, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

In July, the Northern Hemisphere is experiencing its summer season because the North Pole is now tilted towards the sun (**Figure 7m-3**). Some conspicuous hot-spots include the south-central United States, Arizona and northwest Mexico, northern Africa, the Middle East, India, Pakistan, and Afghanistan. Temperatures over oceans tend to be relatively cooler because of the land's ability to heat quickly. Two terrestrial areas of cooler temperatures include Greenland and the Plateau of Tibet. In these regions, most of the incoming solar radiation is sent back to space because of the presence of reflective ice and snow.

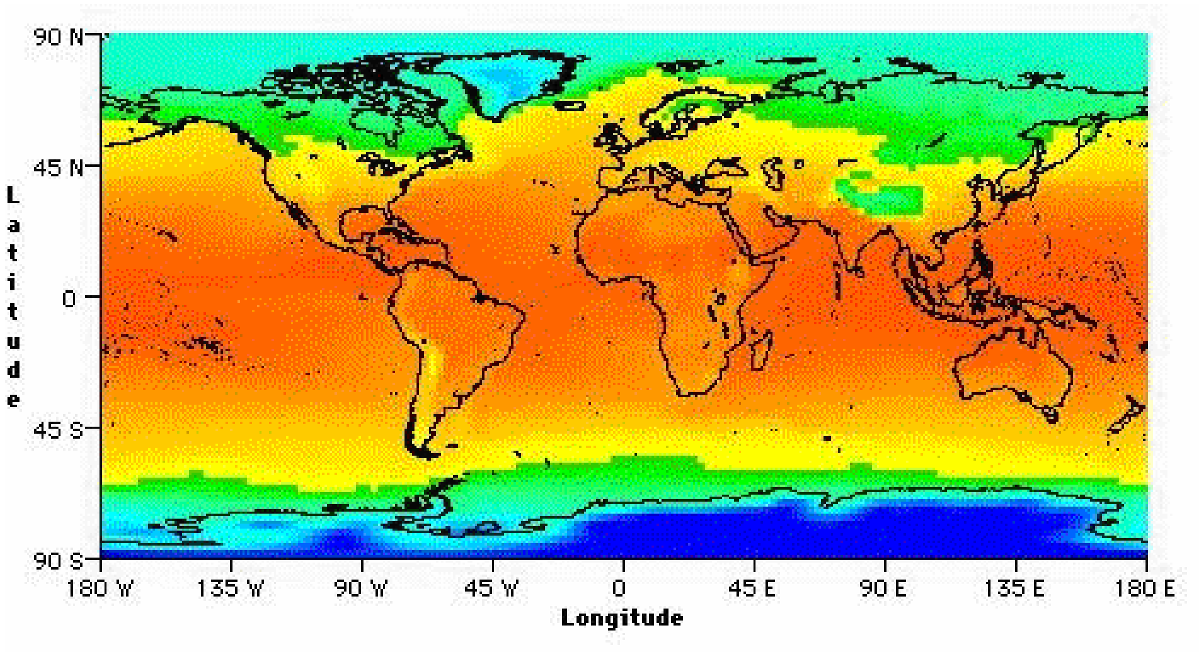
In the Southern Hemisphere, temperatures over the major landmasses are generally cooler than ocean surfaces at the same latitude (**Figure 7m-3**). Antarctica is bitterly cold because it is experiencing total darkness. Note that Antarctica is much colder than the Arctic was during its winter season (**Figures 7m-2** and **7m-3**). The Arctic consists mainly of ocean. During the summer, this surface is able to absorb considerable quantities of sunlight which is then converted into heat energy. The heat stored in the ocean is carried over into the winter season. Antarctica has a surface composed primarily of snow and ice. This surface absorbs only a small amount of the solar radiation during the summer. So it never really heats up. As a result, the amount of heat energy stored into the winter season is minimal.



{PRIVATE}**Figure 7m-3:** Mean July air temperature for the Earth's surface, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

Figure 7m-4 describes average ***annual global temperature*** data for the Earth for the period 1982-1994. The patterns of temperature distribution on this figure are once again mostly latitudinal. However, the latitudinal banding is partially upset by the fact that water bodies are generally warmer than land surfaces. The image also shows the effect of altitude (e.g., Himalayas and Andes mountains) and ***albedo*** (Greenland and Antarctica) on surface air temperature.

Average Annual Global Temperature 1982-1994



Temperature Scale in Kelvin

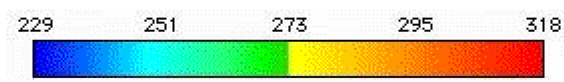


Figure 7m-4: Average annual temperatures for the Earth's surface (1982-94).

(Image generated by *WorldWatcher* software).

(n) Forces Acting to Create Wind

{PRIVATE**}Introduction**

Wind can be defined simply as air in motion. This motion can be in any direction, but in most cases the horizontal component of wind flow greatly exceeds the flow that occurs vertically. The speed of wind varies from absolute calm to speeds as high as 380 kilometers per hour (Mt. Washington, New Hampshire, April 12, 1934). In 1894, strong winds in Nebraska pushed six fully loaded coal cars over 160 kilometers in just over three hours. Over short periods of time surface winds can be quite variable.

Wind develops as a result of spatial differences in ***atmospheric pressure***. Generally, these differences occur because of uneven absorption of ***solar radiation*** at the Earth's surface (**Figure 7n-1**). Wind speed tends to be at its greatest during the daytime when the greatest spatial extremes in atmospheric temperature and pressure exist.

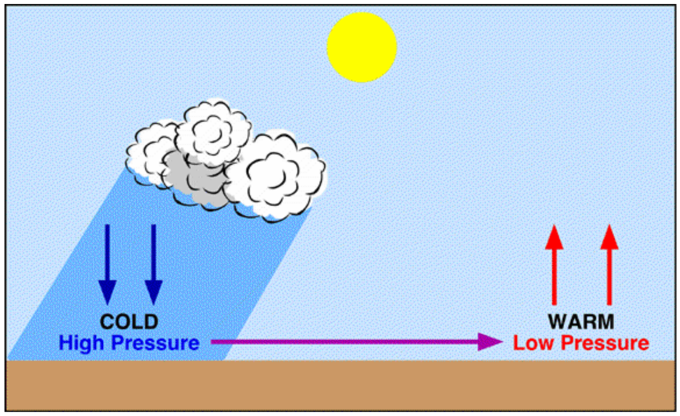


Figure 7n-1: Formation of wind as a result of localized temperature differences.

Wind is often described by two characteristics: wind speed and wind direction. Wind speed is the velocity attained by a mass of air traveling horizontally through the atmosphere. Wind speed is often measured with an ***anemometer*** in kilometers per hour (kmph), miles per hour (mph), knots, or meters per second (mps) (**Figure 7n-2**). Wind direction is measured as the direction from where a wind comes from. For example, a southerly wind comes from the south and blows to the north. Direction is measured by an instrument called a ***wind*** (**Figure 7n-2**). Both of these instruments are positioned in the atmospheric environment at a standard distance of 10 meters above the ground surface.

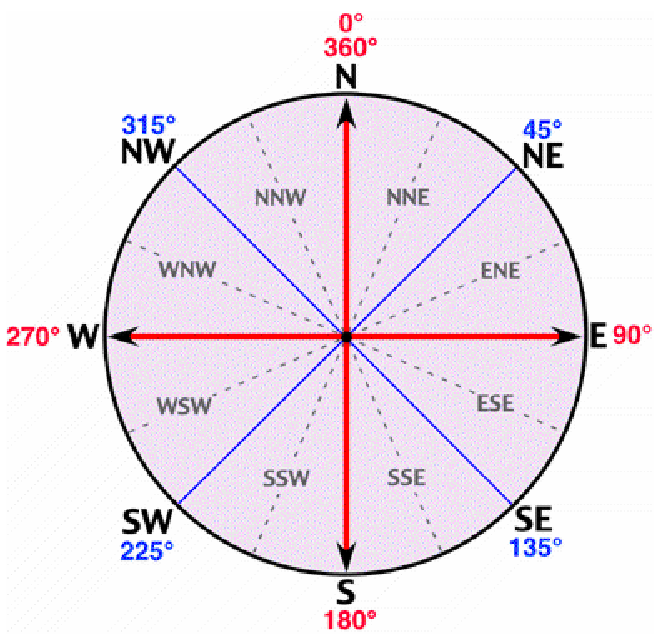
Wind speed can also be measured without the aid of instruments using the ***Beaufort wind scale*** (**Table 7n-1**). This descriptive scale was originally developed by Admiral Beaufort of the British Navy in the first decade of the 17th century. The purpose for this system was to allow mariners to determine wind speed from simple observations. The Beaufort system has undergone several modifications to standardize its measurement scale and to allow for its use on land. Users of this scale look for specific effects of the wind on the environment to determine speed.

{PRIVATE**}Figure 7n-2:** Meteorological instruments used to measure wind speed and direction. Wind speed is commonly measured with an ***anemometer***. An anemometer consists of three open cups attached to a rotating spindle.
The speed of roation is then converted into a measurement of wind speed. Wind direction is measured with a ***wind***. On the photograph above, the wind vane instrument has a bullet shaped nose attached to a finned tail by a metal bar.

Table 7n-1: Beaufort wind speed scale.

{PRIVATE}**Beaufort Code**	**Speed Miles per Hour**	**Speed Kilometers per Hour**	**Description**	**Effects on the Environment**
0	< 1	< 1	calm	smoke rises vertically
1	2 - 3	1 - 5	light air	smoke drifts slowly
2	4 - 7	6 - 11	light breeze	leaves rustle, wind can be felt, wind vanes move
3	8 - 12	12 - 19	gentle breeze	leaves and twigs on trees move
4	13 - 18	20 - 29	moderate breeze	small tree branches move, dust is picked up from the ground surface
5	19 - 24	30 - 38	fresh breeze	small trees move
6	25 - 31	39 - 51	strong breeze	large branches move, telephone and power overhead wires whistle
7	32 - 38	51 - 61	near gale	trees move, difficult to walk in the wind
8	39 - 46	62 - 74	gale	twigs break off from trees
9	47 - 54	75 - 86	strong gale	branches break off from trees, shingles blown off roofs
10	55 - 63	87 - 101	whole gale	trees become uprooted, structural damage on buildings
11	64 - 74	102 - 120	storm	widespread damage to buildings and trees
12	> 75	> 120	hurricane	severe damage to buildings and trees

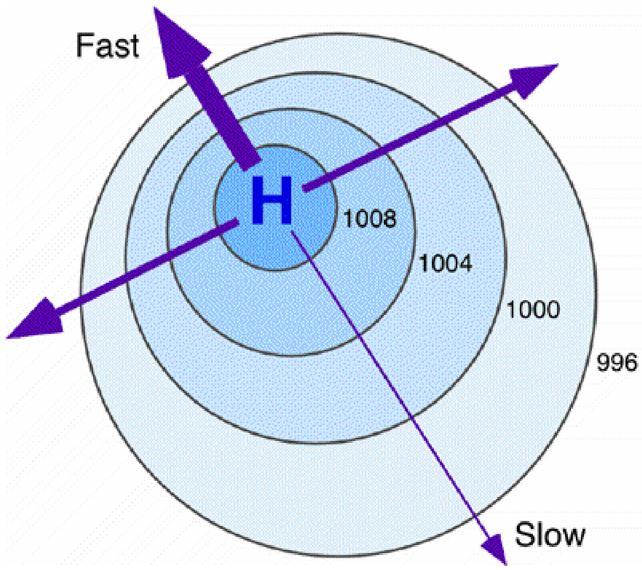
Winds are named according to the compass direction of their source. Thus, a wind from the north blowing toward the south is called a northerly wind. **Figure 7n-3** describes the sixteen principal bearings of wind direction. Most meterological observations report wind direction using one of these sixteen bearings.



{PRIVATE}**Figure 7n-3:** Wind compass describing the sixteen principal bearings used to measure wind direction. This compass is based on the 360 degrees found in a circle.

Horizontally, at the Earth's surface wind always blows from areas of ***high pressure*** to areas of ***low pressure*** (vertically, winds move from areas of low pressure to areas of high pressure), usually at speeds determined by the rate of air pressure change between pressure centers. This situation is comparable to someone skiing down a hill. The skier will of course move from the top of the hill to the bottom of the hill, with the speed of their descent controlled by the gradient or steepness of the slope. Likewise, wind speed is a function of the steepness or gradient of atmospheric air pressure found between high and low pressure systems. When expressed scientifically, pressure change over a unit distance is called ***pressure gradient force***, and the greater this force the faster the winds will blow.

On weather maps, pressure is indicated by drawing ***isolines*** of pressure, called ***isobars***, at regular 4 ***millibar*** intervals (e.g., 996 mb, 1000 mb, 1004 mb, etc.). If the isobars are closely spaced, we can expect the pressure gradient force to be great, and wind speed to be high (see **Figure 7n-4**). In areas where the isobars are spaced widely apart, the pressure gradient is low and light winds normally exist. High speed winds develop in areas where isobars are closer.



{PRIVATE**}Figure 7n-4:** Association between wind speed and distance between isobars. In the illustration above thicker arrows represent relatively faster winds.

Driving Forces

To better understand wind we must recognize that it is the result of a limited number of accelerating and decelerating forces, and that the action of these forces is controlled by specific fundamental natural laws. Sir Isaac Newton formulated these laws as several laws of motion. The first law suggests that an object that is stationary will remain stationary, and an object in motion will stay in motion as long as no opposing force is put on the object. As a result of this law, a puck sent in flight from a blade of a hockey stick will remain in motion until friction slows it down or the goalie makes a save. This law also suggests that once in motion an object's path should be straight.

Newton's second law of motion suggests that the ***force*** put on an object equals its ***mass*** multiplied by the acceleration produced. The term force in this law refers to the total or net effect of all the forces acting on an object. Mathematically, this law is written as:

Force = Mass x Acceleration or Acceleration = Force/Mass

From this natural law of motion we can see that the acceleration of an object is directly proportional to the net force pushing or pulling that body and inversely proportional to the mass of the body. Thus, the greater the force created by the movement of a hockey player's stick the faster the puck will travel. This law also suggests that if the player used a larger (more massive) puck more force would have to be applied to it to get it to travel as fast as a less massive puck.

In the previous lecture we briefly examined one of the forces, ***pressure gradient force***, acting on wind. Let us return to this force and examine it in greater detail and in relation to Newton's laws of motion. We will also examine the effects of three other forces that act on air in motion.

Pressure gradient force is the primary force influencing the formation of wind from local to global scales. This force is determined by the spatial pattern of atmospheric pressure at any given moment in time. **Figure 7n-5** illustrates two different pressure gradient scenarios and their relative effect on wind speed.

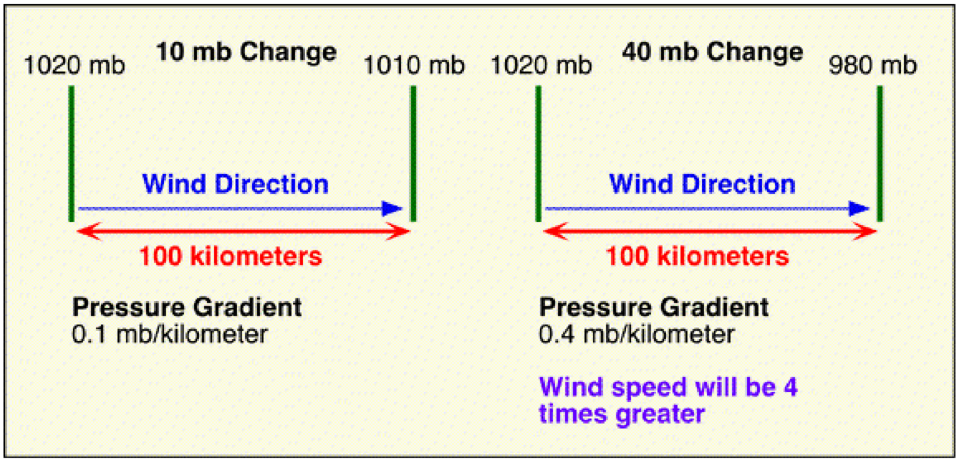


Figure 7n-5: Effect of pressure gradient on wind speed.

The two diagrams display the relative relationship between pressure gradient and wind speed. This relationship is linear and positive. As a result, quadrupling the pressure gradient increases wind speed by a factor of four. This is what we would expect according to Newton's second law of motion, assuming the mass of the wind is unchanged.

We can also describe pressure gradient acceleration mathematically with the following equation:

$$F(ms^{-1}) = \left| \frac{1}{D} \bullet \left(\frac{P_1 - P_2}{n} \right) \right|$$

where:

D = ***density*** of air (average density of surface air is 1.29 kilograms per cubic meter)

P2 = pressure at point 2 in ***newtons***

P1 = pressure at point 1 in ***newtons***

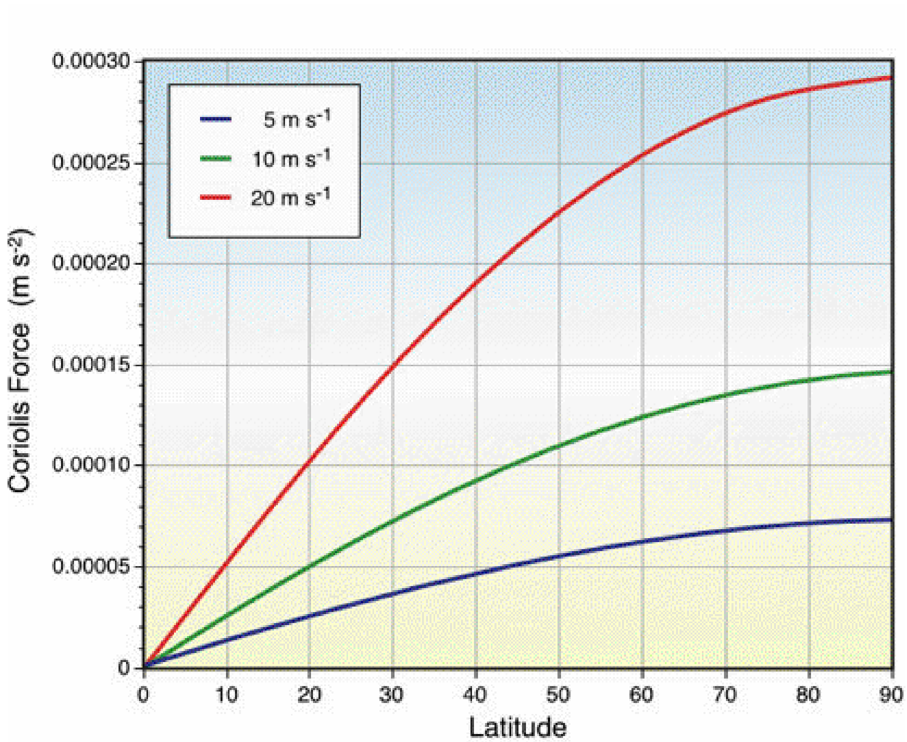
n = distance between the two points in meters

From this equation we can determine the speed of moving air between two points in meters per second by knowing three variables: the ***density*** of the moving air; the change in ***pressure*** between the points of interest in ***newtons***; and the distance between the two points in meters. For example, to determine the wind speed between two points for moving air with a density of 1.29 kilograms per cubic meter, a pressure difference of 400 newtons, and a distance of 300,000 meters, the following calculations would be performed:

$$F(ms^{-1}) = \left| \frac{1}{1.29} \bullet \left(\frac{400}{300,000} \right) \right| = 0.00103 ms^{-1}$$

With the calculator below, try your hand at the calculations above.

The rotation of the Earth creates another force, termed ***Coriolis force***, which acts upon wind and other objects in motion in very predictable ways. According to Newton's first law of motion, air will remain moving in a straight line unless it is influenced by an unbalancing force. The consequence of Coriolis force opposing pressure gradient acceleration is that the moving air changes direction. Instead of wind blowing directly from high to low pressure, the rotation of the Earth causes wind to be deflected off course. In the Northern Hemisphere, wind is deflected to the right of its path, while in the Southern Hemisphere it is deflected to the left. The magnitude of the Coriolis force varies with the velocity and the latitude of the object (see **Figure 7n-6**). Coriolis force is absent at the equator, and its strength increases as one approaches either pole. Furthermore, an increase in wind speed also results in a stronger Coriolis force, and thus in greater deflection of the wind. Coriolis force only acts on air when it has been sent into motion by pressure gradient force. Finally, Coriolis force only influences wind direction and never wind speed.



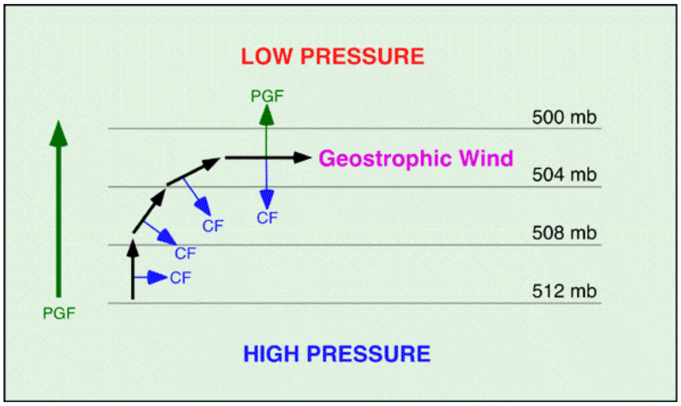
{PRIVATE}**Figure 7n-6:** The strength of Coriolis force is influenced by latitude and the speed of the moving object.

Centripetal acceleration is the third force that can act on moving air. It acts only on air that is flowing around centers of circulation. Centripetal acceleration is also another force that can influence the direction of wind. Centripetal acceleration creates a force directed at right angles to the flow of the wind and inwards towards the centers of rotation (e.g., low and high pressure centers). This force produces a circular pattern of flow around centers of high and low pressure. Centripetal acceleration is much more important for circulations smaller than the ***mid-latitude cyclone***.

The last force that can influence moving air is ***frictional deceleration***. Friction can exert an influence on wind only after the air is in motion. Frictional drag acts in a direction opposite to the path of motion causing the moving air to decelerate (see Newton's first and second laws of motion). Frictional effects are limited to the lower one kilometer above the Earth's surface.

Geostrophic Wind

Air under the influence of both the ***pressure gradient force*** and ***Coriolis force*** tends to move parallel to ***isobars*** in conditions where friction is low (1000 meters above the surface of the Earth) and isobars are straight. Winds of this type are usually called ***geostrophic winds***. Geostrophic winds come about because pressure gradient force and Coriolis force come into balance after the air begins to move (**Figure 7n-7**).



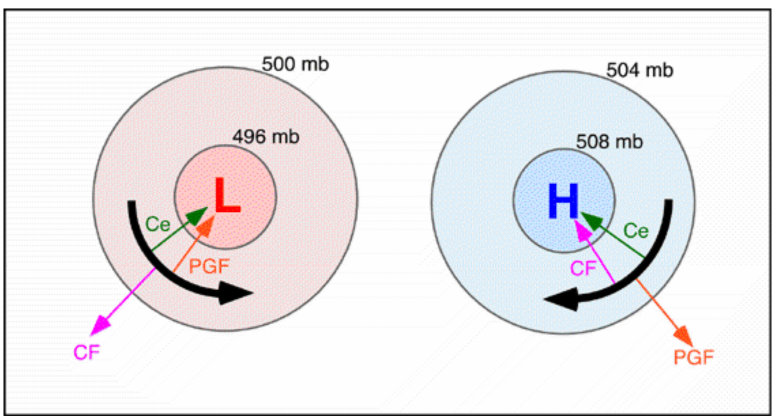
{PRIVATE}Figure 7n-7: A geostrophic wind flows parallel to the isobars. In this model of wind flow in the Northern Hemisphere, wind begins as a flow of air perpendicular to the isobars (measured in ***millibars***) under the primary influence of the pressure gradient force (**PGF**). As the movement begins, the Coriolis force (**CF**) begins to influence the moving air causing it to deflect to the right of its path. This deflection continues until the pressure gradient force and Coriolis force are opposite and in balance with each other.

Figure 7n-7 models air flow in the Northern Hemisphere. In the Southern Hemisphere, Coriolis acceleration acts on moving air by deflecting it to the left instead of the right.

Finally, **Buy Ballot's Law** states that when you stand with your back to a geostrophic wind in the Northern Hemisphere the center of low pressure will be to your left and the high pressure to your right. The opposite is true for the Southern Hemisphere.

Gradient Wind

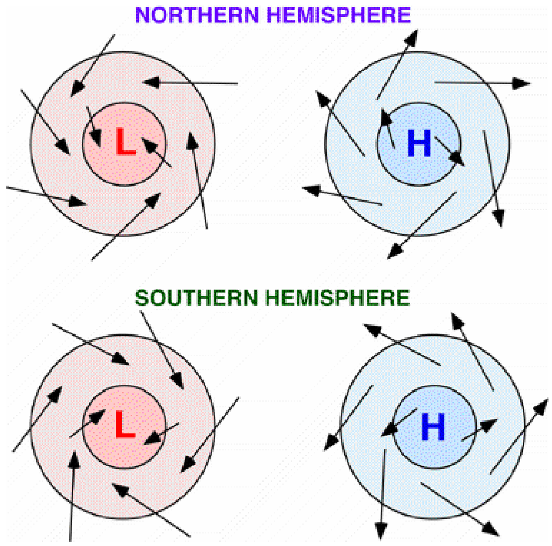
Wind above the Earth's surface does not always travel in straight lines. In many cases winds flow around the curved ***isobars*** of a high (***anticyclone***) or low (***cyclone***) pressure center. A wind that blows around curved isobars above the level of friction is called a ***gradient wind***. Gradient winds are slightly more complex than geostrophic winds because they include the action of yet another physical force. This force is known as ***centripetal force*** and it is always directed toward the center of rotation. The following **figure** describes the forces that produce gradient winds around high and low pressure centers (**Figure 7n-8**). Around a low, the gradient wind consists of the ***pressure gradient force*** and ***centripetal force*** acting toward the center of rotation, while ***Coriolis force*** acts away from the center of the low. In a high pressure center, the Coriolis and centripetal forces are directed toward the center of the high, while the pressure gradient force is directed outward.



{PRIVATE}Figure 7n-8: The balance of forces that create a gradient wind in the Northern Hemisphere (**PFG** = pressure gradient force; **CF** = Coriolis force; **Ce =** centripetal force). In this diagram, **CF** = **Ce** + **PFG** for the low, and **PFG** = **CF** + **Ce** for the high.

Friction Layer Wind

Surface winds on a weather map do not blow exactly parallel to the isobars as in geostrophic and gradient winds. Instead, surface winds tend to cross the isobars at an angle varying from 10 to 45 degrees. Close to the Earth's surface, ***friction*** reduces the wind speed, which in turn reduces the ***Coriolis force***. As a result, the reduced Coriolis force no longer balances the ***pressure gradient force***, and the wind blows across the isobars toward or away from the pressure center. The pressure gradient force is now balanced by the sum of the ***frictional force*** and the Coriolis force. Thus, we find surface winds blowing counterclockwise and inward into a surface low, and clockwise and out of a surface high in the Northern Hemisphere. In the Southern Hemisphere, the Coriolis force acts to the left rather than the right. This causes the winds of the Southern Hemisphere to blow clockwise and inward around surface lows, and counterclockwise and outward around surface highs (see **Figure 7n-9** below).



{PRIVATE}**Figure 7n-9:** Circulation patterns of high and low pressure systems in the North and South Hemisphere.

(o) Local and Regional Wind Systems

{PRIVATE}**Thermal Circulations**

As discussed earlier, winds blow because of differences in ***atmospheric pressure***. ***Pressure gradients*** may develop on a local to a global scale because of differences in the heating and cooling of the Earth's surface. Heating and cooling cycles that develop daily or annually can create several common local or regional ***thermal wind systems***. The basic circulation system that develops is described in the generic **illustrations** below.

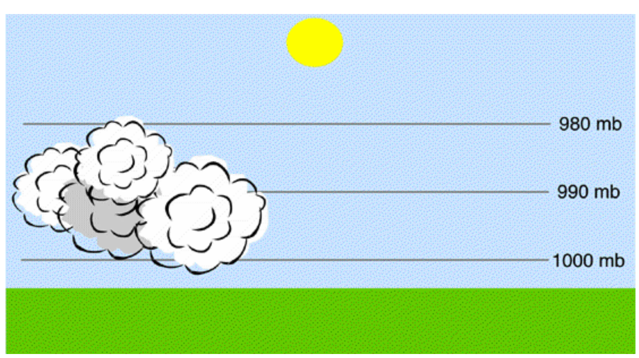


Figure 7o-1: Cross-section of the atmosphere with uniform horizontal atmospheric pressure.

In this first **diagram** (**Figure 7o-1**), there is no horizontal temperature or pressure gradient and therefore no wind. Atmospheric pressure decreases with altitude as depicted by the drawn ***isobars*** (1000 to 980 ***millibars***). In the second **diagram** (**Figure 7o-2**), the potential for solar heating is added which creates contrasting surface areas of temperature and ***atmospheric pressure***. The area to the right receives more ***solar radiation*** and the air begins to warm from heat energy transferred from the ground through ***conduction*** and ***convection***. The vertical distance between the isobars becomes greater as the air rises. To the far left, less radiation is received because of the presence of cloud, and this area becomes relatively cooler than the area to the right. In the upper atmosphere, a ***pressure gradient*** begins to form because of the rising air and upward spreading of the isobars. The air then begins to flow in the upper atmosphere from ***high pressure*** to ***low pressure***.

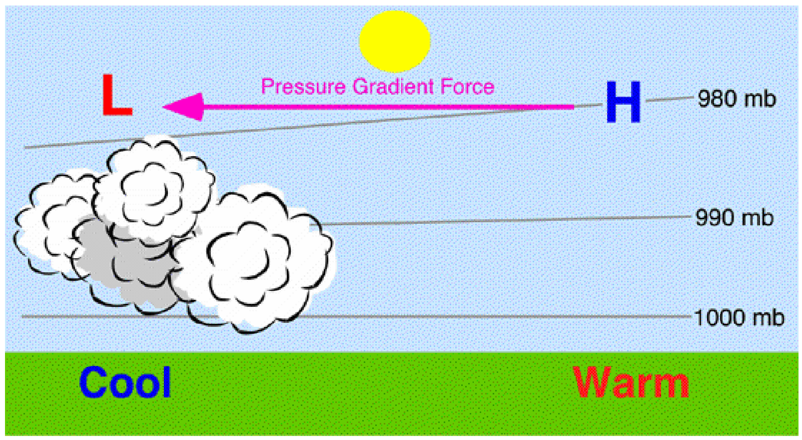


Figure 7o-2: Development of air flow in the upper atmosphere because of surface heating.

Figure 7o-3 shows the full circulation system in action. Beneath the upper atmosphere high is a ***thermal low*** pressure center created from the heating of the ground surface. Below the upper atmosphere low is a ***thermal high*** created by the relatively cooler air temperatures and the descend air from above. Surface air temperatures are cooler here because of the obstruction of shortwave radiation absorption at the Earth's surface by the cloud. At the surface, the wind blows from the high to the low pressure. Once at the low, the wind rises up to the upper air high pressure system because of thermal buoyancy and outflow in the upper atmosphere. From the upper high, the air then travels to the upper air low, and then back down to the surface high to complete the circulation cell. The circulation cell is a closed system that redistributes air in an equitable manner. It is driven by the greater heating of the surface air in the right of the diagram.

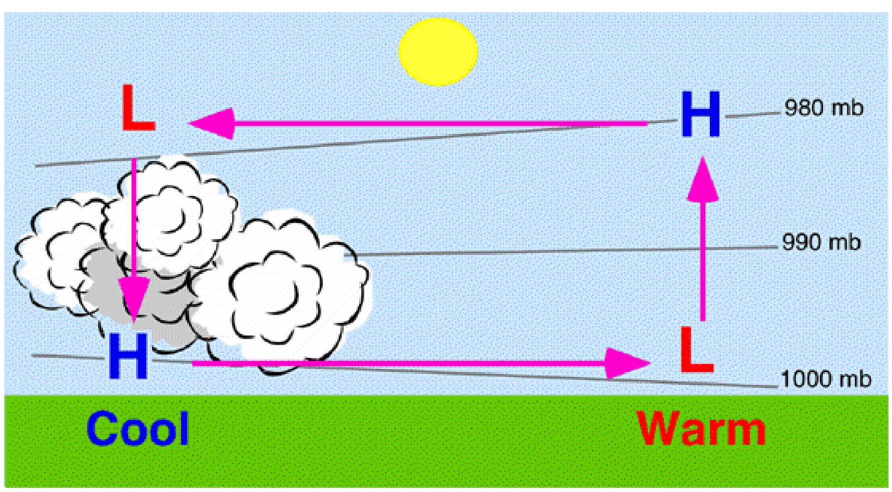


Figure 7o-3: Development of a closed atmospheric circulation cell because of surface heating.

Sea and Land Breezes

Sea and ***land breezes*** are types of ***thermal circulation*** systems that develop at the interface of land and ocean. At this interface, the dissimilar heating and cooling characteristics of land and water initiate the development of an atmospheric pressure gradient which causes the air in these areas to flow.

During the daytime land heats up much faster than water as it receives solar radiation from the sun (**Figure 7o-4**). The warmer air over the land then begins to expand and rise forming a thermal low. At the same time, the air over the ocean becomes a cool high because of water's slower rate of heating. Air begins to flow as soon as there is a significant difference in air temperature and pressure across the land to sea gradient. The development of this pressure gradient causes the heavier cooler air over the ocean to move toward the land and to replace the air rising in the thermal low. This localized air flow system is called a ***sea breeze***. Sea breeze usually begins in midmorning and reaches its maximum strength in the later afternoon when the greatest temperature and pressure contrasts exist. It dies down at sunset when air temperature and pressure once again become similar across the two surfaces.

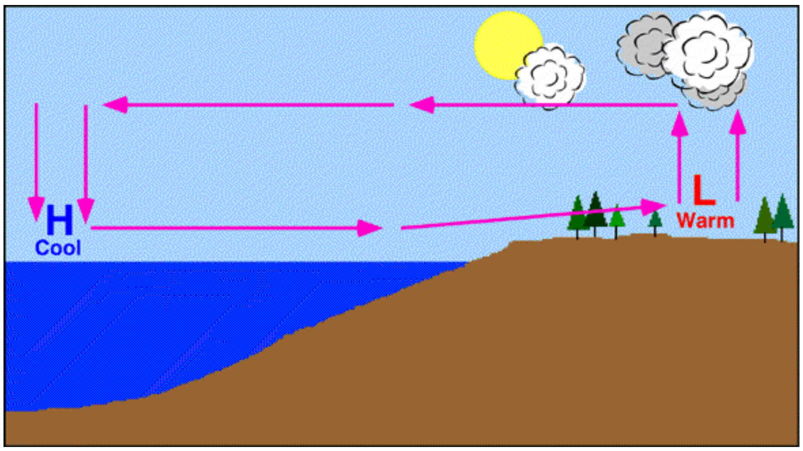


Figure 7o-4: Daytime development of sea breeze.

At sunset, the land surface stops receiving radiation from the sun (**Figure 7o-5**). As night continues the land surface begins losing heat energy at a much faster rate than the water surface. After a few hours, air temperature and pressure contrasts begin to develop between the land and ocean surfaces. The land surface being cooler than the water becomes a thermal high pressure area. The ocean becomes a warm thermal low. Wind flow now moves from the land to the open ocean. This type of localized air flow is called a ***land breeze***.

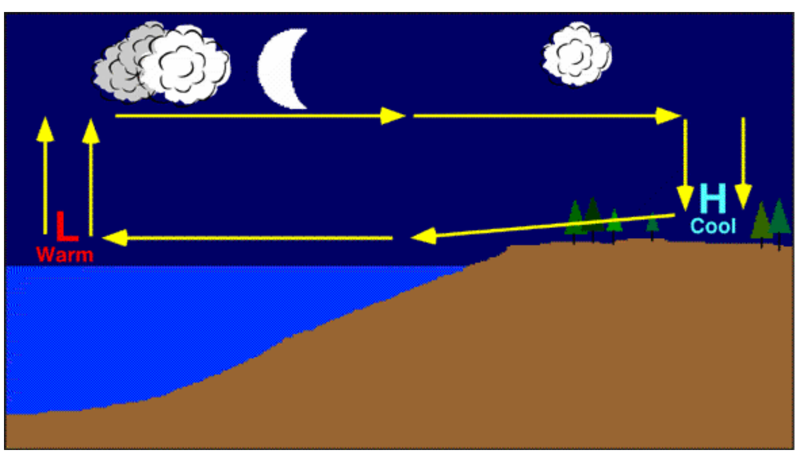


Figure 7o-5: Nighttime development of land breeze.

Mountain and Valley Breezes

Mountain and ***valley breezes*** are common in regions with great topographic relief (**Figure 7o-6** and **7o-7**). A ***valley breeze*** develops during the day as the sun heats the land surface and air at the valley bottom and sides (**Figure 7o-6**). As the air heats it becomes less dense and buoyant and begins to flow gently up the valley sides. Vertical ascent of the air rising along the sides of the mountain is usually limited by the presence of a ***temperature inversion*** layer. When the ascending air currents encounter the inversion they are forced to move horizontally and then back down to the valley floor. This creates a self-contained circulation system. If conditions are right, the rising air can condense and form into cumuliform clouds.

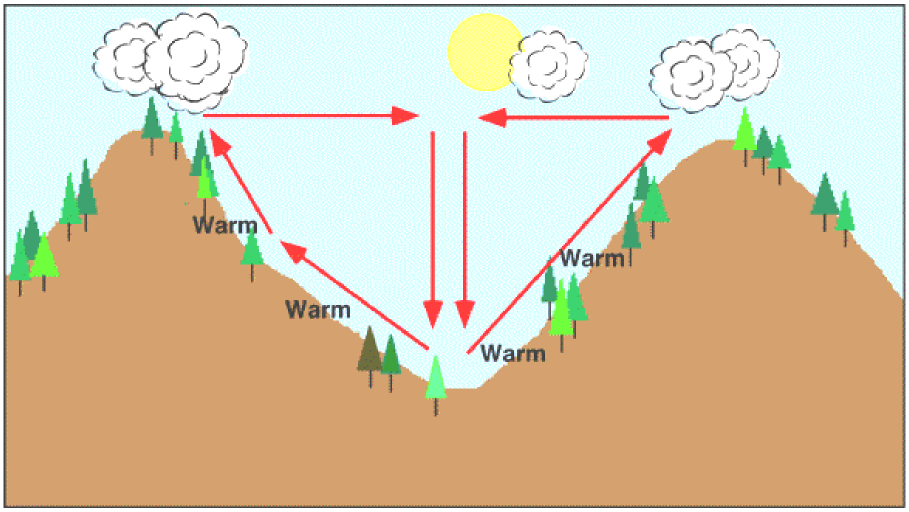


Figure 7o-6: Daytime development of valley breeze.

During the night, the air along the mountain slopes begins to cool quickly because of ***longwave radiation*** loss (**Figure 7o-7**). As the air cools, it becomes more dense and begins to flow downslope causing a ***mountain breeze***. Convergence of the draining air occurs at the valley floor and forces the air to move vertically upward. The upward movement is usually limited by the presence of a ***temperature inversion*** which forces the air to begin moving horizontally. This horizontal movement completes the circulation cell system. In narrowing terrain, mountain winds can accelerate in speed because of the ***venturi*** effect. Such winds can attain speeds as high has 150 kilometers per hour.

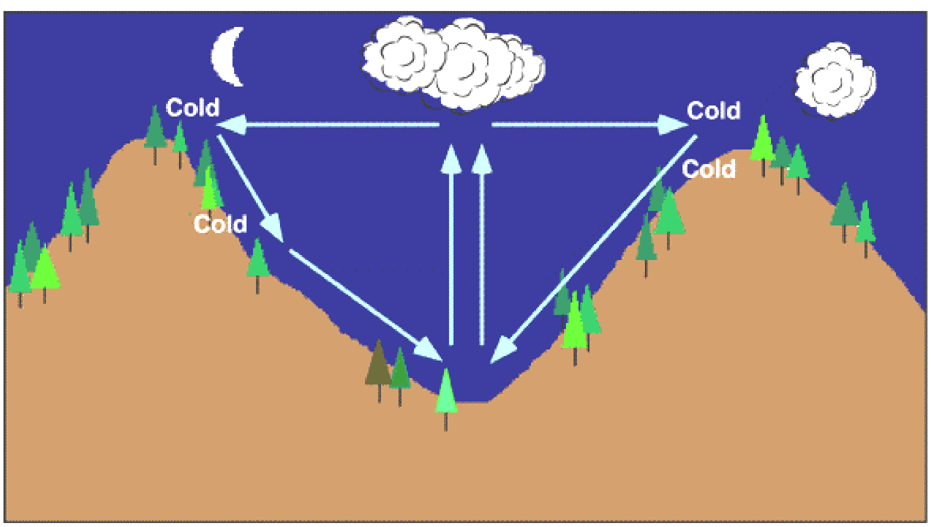


Figure 7o-7: Nightime development of mountain breeze.

Monsoon Winds

Monsoons are regional scale wind systems that predictably change direction with the passing of the seasons. Like land/sea breezes, these wind systems are created by the temperature contrasts that exist between the surfaces of land and ocean. However, monsoons are different from land/sea breezes both spatially and temporally. Monsoons occur over distances of thousands of kilometers, and their two dominant patterns of wind flow act over an annual time scale.

During the summer, monsoon winds blow from the cooler ocean surfaces onto the warmer continents. In the summer, the continents become much warmer than the oceans because of a number of factors. These factors include:

Specific heat differences between land and water.

Greater ***evaporation*** over water surfaces.

Subsurface mixing in ocean basins which redistributes heat energy through a deeper layer.

Precipitation is normally associated with the summer monsoons. Onshore winds blowing inland from the warm ocean are very high in humidity, and slight cooling of these air masses causes condensation and rain. In some cases, this precipitation can be greatly intensified by ***orographic uplift***. Some highland areas in Asia receive more than 10 meters of rain during the summer months.

In the winter, the wind patterns reverse as the ocean surfaces are now warmer. With little solar energy available, the continents begin cooling rapidly as longwave radiation is emitted to space. The ocean surface retains its heat energy longer because of water's high specific heat and subsurface mixing. The winter monsoons bring clear dry weather and winds that flow from land to sea.

Figure 7o-8 illustrates the general wind patterns associated with the winter and summer monsoons in Asia. The Asiatic monsoon is the result of a complex climatic interaction between the distribution of land and water, topography, and tropical and mid-latitudinal circulation. In the summer, a low pressure center forms over northern India and northern Southeast Asia because of higher levels of received solar insolation. Warm moist air is drawn into the thermal lows from air masses over the Indian Ocean. Summer heating also causes the development of a strong latitudinal pressure gradient and the development of an easterly jet stream at an altitude of about 15 kilometers and a latitude of 25 degrees North. The jet stream enhances rainfall in Southeast Asia, in the Arabian Sea, and in South Africa. When autumn returns to Asia the thermal extremes between land and ocean decrease and the westerlies of the mid-latitudes move in. The easterly jet stream is replaced with strong westerly winds in the upper atmosphere. Subsidence from an upper atmosphere cold low above the Himalayas produces outflow that creates a surface high pressure system that dominates the weather in India and Southeast Asia.

Monsoon wind systems also exist in Australia, Africa, South America, and North America.

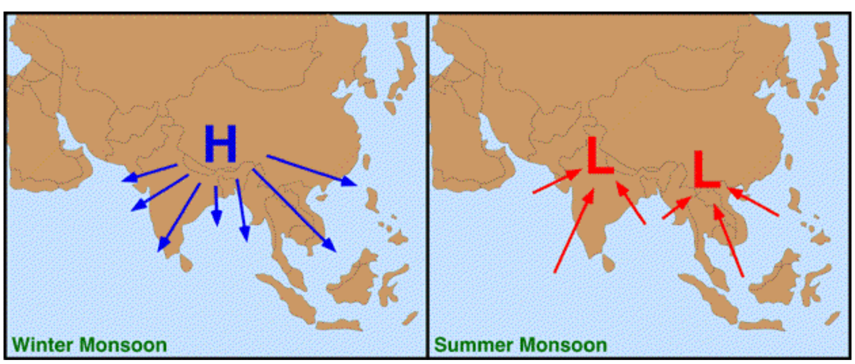


Figure 7o-8: Winter and summer monsoon wind patterns for southeast Asia.

(p) Global Scale Circulation of the Atmosphere

Simple Model of Global Circulation

We can gain an understanding of how global circulation works by developing two simplified graphical models of processes that produce this system. The first model will be founded on the following simplifying assumptions:

The Earth is not rotating in space.

The Earth's surface is composed of similar materials.

The global reception of ***solar insolation*** and loss of ***longwave radiation*** cause a temperature gradient of hotter air at the equator and colder air at the poles.

Based on these assumptions, air circulation on the Earth should approximate the patterns shown on **Figure 7p-1**. In this illustration, each hemisphere contains one three-dimensional circulation cell.

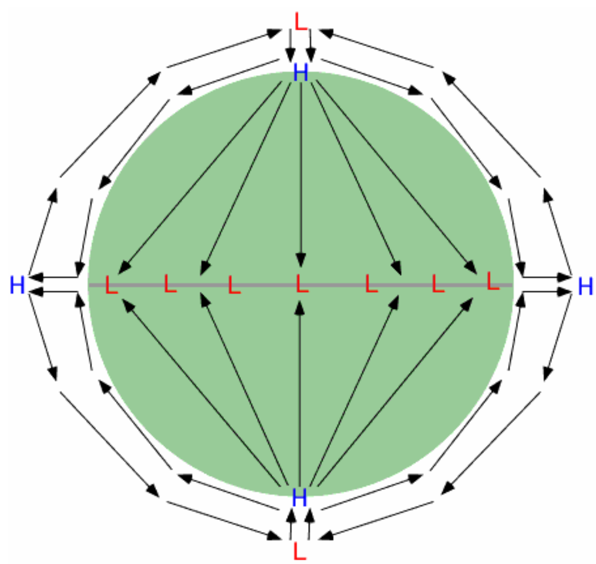


Figure 7p-1: Simplified one-cell global air circulation patterns.

As described in the diagram above, surface air flow is from the poles to the equator. When the air reaches the equator, it is lifted vertically by the processes of ***convection*** and ***convergence***. When it reaches the top of the troposphere, it begins to flow once again horizontally. However, the direction of flow is now from the equator to the poles. At the poles, the air in the upper atmosphere then descends to the Earth's surface to complete the cycle of flow.

Three Cell Model of Global Circulation

If we eliminate the first assumption, the pattern of flow described in the model above would be altered, and the mesoscale flow of the atmosphere would more closely approximate the actual global circulation on the Earth. Planetary rotation would cause the development of three circulation cells in each hemisphere rather than one (see **Figure 7p-2**). These three circulation cells are known as the: ***Hadley cell***; ***Ferrel cell***; and ***Polar cell***.

In the new model, the equator still remains the warmest location on the Earth. This area of greater heat acts as zone of ***thermal lows*** known as the ***intertropical convergence zone*** (**ITCZ**). The Intertropical Convergence Zone draws in surface air from the subtropics. When this subtropical air reaches the equator, it rises into the upper atmosphere because of ***convergence*** and ***convection***. It attains a maximum vertical altitude of about 14 kilometers (top of the ***troposphere***), and then begins flowing horizontally to the North and South Poles. ***Coriolis force*** causes the deflection of this moving air in the upper atmosphere, and by about 30 degrees of latitude the air begins to flow zonally from west to east. This ***zonal*** flow is known as the ***subtropical jet stream.*** The zonal flow also causes the accumulation of air in the upper atmosphere as it is no longer flowing ***meridionally***. To compensate for this accumulation, some of the air in the upper atmosphere sinks back to the surface creating the ***subtropical high pressure zone***. From this zone, the surface air travels in two directions. A portion of the air moves back toward the equator completing the circulation system known as the ***Hadley cell***. This moving air is also deflected by the Coriolis effect to create the ***Northeast Trades*** (right deflection) and ***Southeast Trades*** (left deflection). The surface air moving towards the poles from the subtropical high zone is also deflected by Coriolis acceleration producing the ***Westerlies***. Between the latitudes of 30 to 60 degrees North and South, upper air winds blow generally towards the poles. Once again, Coriolis force deflects this wind to cause it to flow west to east forming the ***polar jet stream*** at roughly 60 degrees North and South. On the Earth's surface at 60 degrees North and South latitude, the subtropical Westerlies collide with cold air traveling from the poles. This collision results in ***frontal uplift*** and the creation of the ***subpolar lows*** or ***mid-latitude cyclones***. A small portion of this lifted air is sent back into the ***Ferrel cell*** after it reaches the top of the troposphere. Most of this lifted air is directed to the ***polar vortex*** where it moves downward to create the ***polar high***.

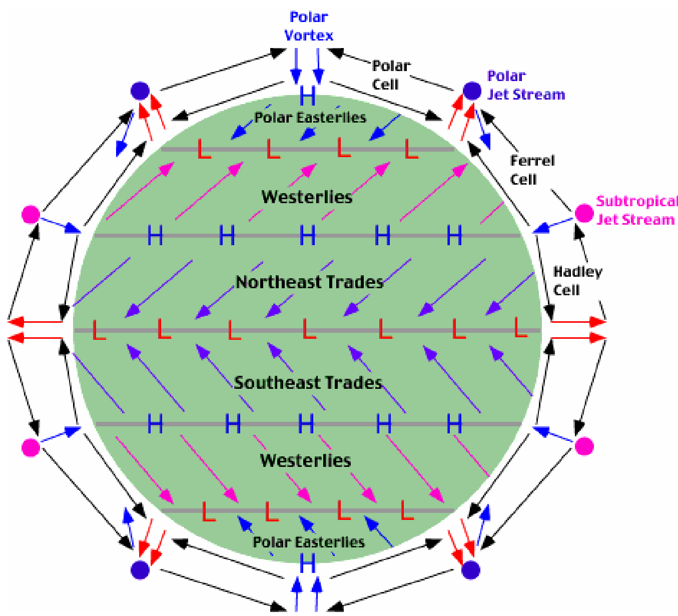


Figure 7p-2: Simplified global three-cell surface and upper air circulation patterns.

Actual Global Surface Circulation

Figure 7p-3 describes the actual surface circulation for the Earth as determined from 39 years of record. Average circulation patterns can be viewed on this animation for every month of the year. The circulation patterns seen differ somewhat from the three cell model in **Figure 7p-2**. These differences are caused primarily by two factors. First, the Earth's surface is not composed of uniform materials. The two surface materials that dominate are water and land. These two materials behave differently in terms of heating and cooling causing latitudinal pressure zones to be less uniform. The second factor influencing actual circulation patterns is elevation. Elevation tends to cause pressure centers to become intensified when altitude is increased. This is especially true of high pressure systems.

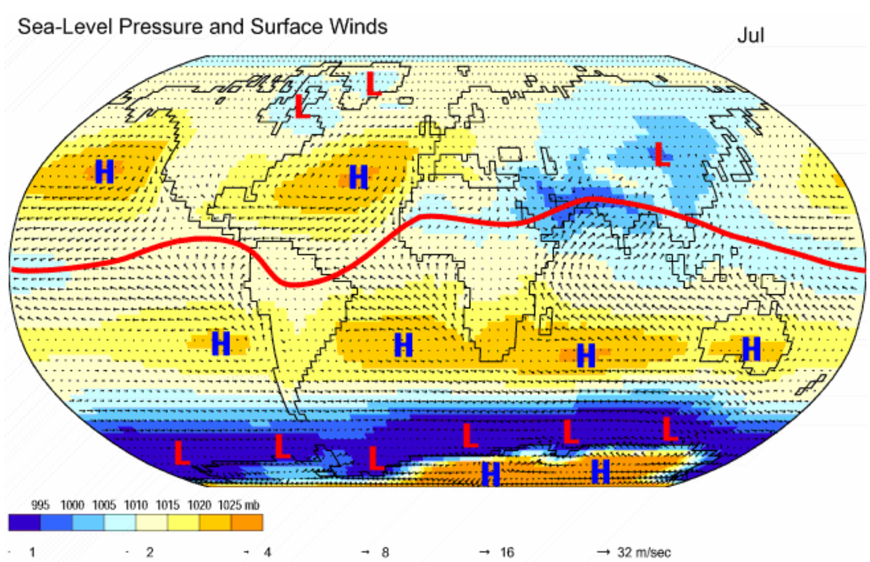
{PRIVATE**}Figure 7p-3:** Monthly average sea-level pressure and prevailing winds for the Earth's surface, 1959-1997. Atmosphere pressure values are adjusted for elevation and are described relative to sea-level. The slider at the bottom of the image allows you change the time of month. Pressure values are indicated by color shading (see the legend in the graphic). Blue shades indicate pressure lower than the global average, while yellow to orange shades are higher than average measurements. Wind direction is shown with arrows. Wind speed is indicated by the length of these arrows (see the legend on the graphic). (**Source:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

(To view this animation your browser must have Apple's **QuickTime** plug-in and your monitor should be set for thousands of colors. The **QuickTime** plug-in is available for Macintosh, Windows 95, Windows 98 and Windows NT computers and can be downloaded from the World Wide Web site ***www.apple.com/quicktime***).

Figures 7p-4 and **7p-5** are single frames from the animation above for January and July, respectively. On these modified graphics, we can better visualize the ***intertropical convergence zone*** (**ITCZ**), ***subtropical high pressure zone***, and the ***subpolar lows***. The intertropical convergence zone is identified on the **figures** by a red line. The formation of this band of low pressure is the result of solar heating and the convergence of the trade winds. In January, the intertropical convergence zone is found south of the equator (**Figure 7p-4**). During this time period, the Southern Hemisphere is tilted towards the sun and receives higher inputs of shortwave radiation. Note that the line representing the intertropical convergence zone is not straight and parallel to the lines of latitude. Bends in the line occur because of the different heating characteristics of land and water. Over the continents of Africa, South America, and Australia, these bends are toward the South Pole. This phenomenon occurs because land heats up faster then ocean.

{PRIVATE}**Figure 7p-4:** Mean January prevailing surface winds and centers of atmospheric pressure, 1959-1997. The red line on this image represents the ***intertropical convergence zone*** (**ITCZ**). Centers of high and low pressure have also been labeled. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

During July, the intertropical convergence zone (ITCZ) is generally found north of the equator (**Figure 7p-5**). This shift in position occurs because the altitude of the sun is now higher in the Northern Hemisphere. The greatest spatial shift in the ITCZ, from January to July, occurs in the eastern half of the image. This shift is about 40 degrees of latitude in some places. The more intense July sun causes land areas of Northern Africa and Asia rapidly warm creating the **Asiatic Low** which becomes part of the ITCZ. In the winter months, the intertropical convergence zone is pushed south by the development of an intense high pressure system over central Asia (compare **Figures 7p-4** and **7p-5**). The extreme movement of the ITCZ in this part of the world also helps to intensify the development of a regional winds system called the Asian ***monsoon***.



{PRIVATE}**Figure 7p-5:** Mean July prevailing surface winds and centers of atmospheric pressure, 1959-1997. The red line on this image represents the ***intertropical convergence zone*** (**ITCZ**). Centers of high and low pressure have also been labeled. (**Source of Original Modified Image:** Climate Lab Section of the

Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

The ***subtropical high pressure zone*** does not form a uniform area of high pressure stretching around the world in reality. Instead, the system consists of several localized ***anticyclonic*** cells of high pressure. These systems are located roughly at about 20 to 30 degrees of latitude and are labeled with the letter **H** on **Figures 7p-4** and **7p-5**. The subtropical high pressure systems develop because of the presence of descending air currents from the ***Hadley cell***. These systems intensify over the ocean during the summer or high sun season. During this season, the air over the ocean bodies remains relatively cool because of the slower heating of water relative to land surfaces. Over land, intensification takes place in the winter months. At this time, land cools off quickly, relative to ocean, forming large cold continental ***air masses***.

The ***subpolar lows*** form a continuous zone of low pressure in the Southern Hemisphere at a latitude of between 50 and 70 degrees (**Figures 7p-4** and **7p-5**). The intensity of the subpolar lows varies with season. This zone is most intense during Southern Hemisphere summer (**Figure 7p-4**). At this time, greater differences in temperature exist between air masses found either side of this zone. North of subpolar low belt, summer heating warms subtropical air masses. South of the zone, the ice covered surface of Antarctica reflects much of the ***incoming solar radiation*** back to space. As a consequence, air masses above Antarctica remain cold because very little heating of the ground surface takes place. The meeting of the warm subtropical and cold polar air masses at the subpolar low zone enhances ***frontal uplift*** and the formation of intense low pressure systems.

In the Northern Hemisphere, the subpolar lows do not form a continuous belt circling the globe (**Figures 7p-4** and **7p-5**). Instead, they exist as localized ***cyclonic*** centers of low pressure. In the Northern Hemisphere winter, these pressure centers are intense and located over the oceans just to the south of Greenland and the Aleutin Islands (**Figure 7p-4**). These areas of low pressure are responsible for spawning many ***mid-latitude cyclones***. The development of the subpolar lows in summer only occurs weakly (**Figure 7p-5 -** over Greenland and Baffin Island, Canada), unlike the Southern Hemisphere. The reason for this phenomenon is that considerable heating of the Earth's surface occurs from 60 to 90 degrees North. As a result, cold polar air masses generally do not form.

(q) Upper Air Winds and the Jet Streams

{PRIVATE**}**Winds at the top of the troposphere are generally poleward and westerly in direction. **Figure 7q-1** describes these upper air westerlies along with some other associated weather features. Three zones of westerlies can be seen in each hemisphere on this **illustration**. Each zone is associated with either the ***Hadley***, ***Ferrel***, or ***Polar*** circulation cell.

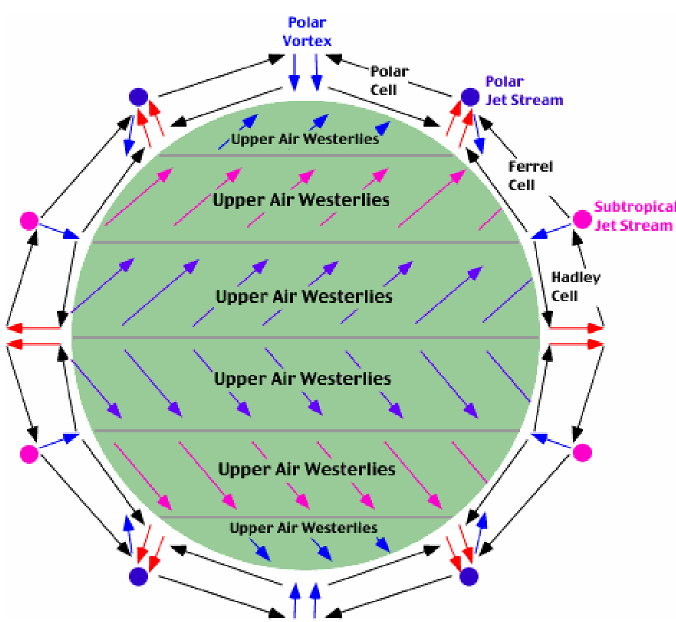


Figure 7q-1: Simplified global three-cell upper air circulation patterns.

The ***polar jet stream*** is formed by the deflection of upper air winds by coriolis acceleration (see **Figure 7q-3** below). It resembles a stream of water moving west to east and has an altitude of about 10 kilometers. Its air flow is intensified by the strong temperature and pressure gradient that develops when cold air from the poles meets warm air from the tropics. Wind velocity is highest in the core of the polar jet stream where speeds can be as high as 300 kilometers per hour. The jet stream core is surrounded by slower moving air that has an average velocity of 130 kilometers per hour in winter and 65 kilometers per hour in summer.

Associated with the polar jet stream is the ***polar front***. The polar front represents the zone where warm air from the subtropics (**pink**) and cold air (**blue**) from the poles meet (see **Figure 7q-3** below). At this zone, massive exchanges of energy occur in the form of storms known as the ***mid-latitude cyclones***. The shape and position of waves in the polar jet stream determine the location and the intensity of the mid-latitude cyclones. In general, mid-latitude cyclones form beneath polar jet stream ***troughs***. The following satellite **image** (**Figure 7q-2**), taken from above the South Pole, shows a number of mid-latitude cyclones circling Antarctica. Each mid-latidude cyclone wave is defined by the cloud development associated with ***frontal uplift***.

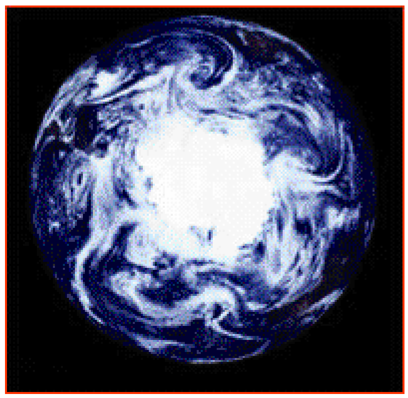


Figure 7q-2: Satellite view of the atmospheric circulation at the South Pole. (**Source:** NASA).

The ***subtropical jet stream*** is located approximately 13 kilometers above the ***subtropical high pressure zone***. The reason for its formation is similar to the polar jet stream. However, the subtropical jet stream is weaker. Its slower wind speeds are the result of a weaker latitudinal temperature and pressure gradient.

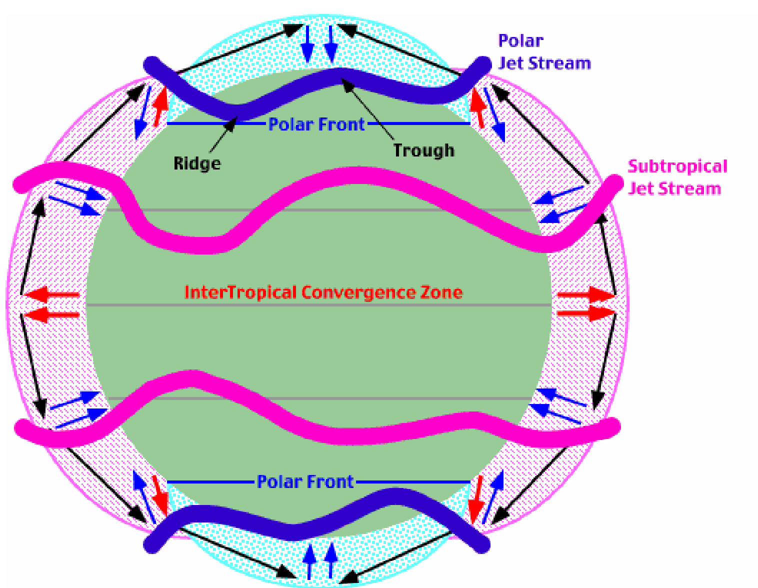
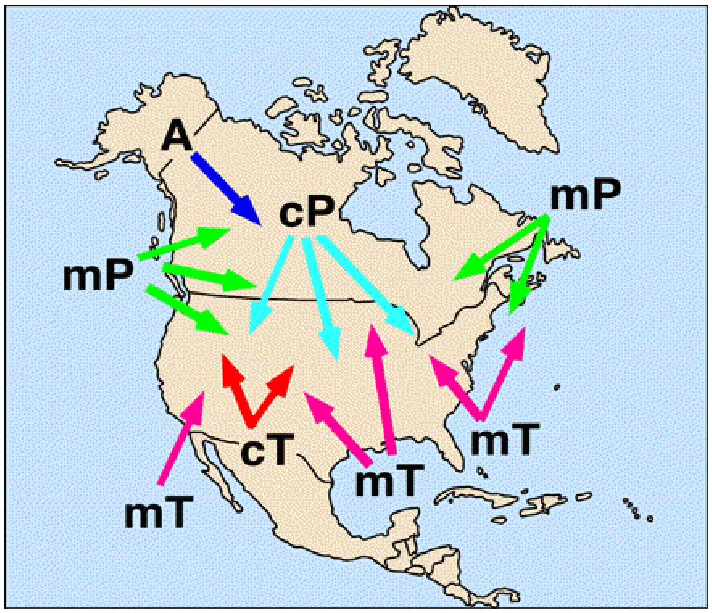


Figure 7q-3: Polar and subtropical jet streams.

(r) Air Masses and Frontal Transitional Zones

{PRIVATE**}**An ***air mass*** is a large body of air of relatively similar temperature and humidity characteristics covering thousands of square kilometers. Typically, air masses are classified according to the characteristics of their ***source region*** or area of formation. A source region can have one of four temperature attributes: **equatorial**, **tropical**, **polar** or **arctic**. Air masses are also classified as being either **continental** or **maritime** in terms of moisture characteristics. Combining these two categories, several possibilities are commonly found associated with North America: ***maritime polar*** (**mP**), ***continental polar*** (**cP**), ***maritime tropical*** (**mT**), ***continental tropical*** (**cT**), and ***continental arctic*** (**A**). The following **diagram** (**Figure 7r-1**) describes the source regions and common patterns of movement for the various types of air masses associated with North America.



{PRIVATE**}Figure 7r-1:** Source sites and movement patterns for North America's major air masses.

Frequently, two air masses, especially in the middle latitudes, develop a sharp boundary or interface, where the temperature difference between them becomes intensified. Such an area of intensification is called a ***frontal zone*** or a ***front***. The boundary between the warm and cold air masses always slopes upwards over the cold air. This is due to the fact that cold air is much denser than warm air. The sloping of warm air over the cold air leads to a forced uplifting (***frontal lifting***) of the warm air if one air mass is moving toward the other. In turn, this uplifting causes condensation to occur and the possibility of precipitation along the frontal boundary.

Frontal zones where the air masses are not moving against each other are called ***stationary fronts***. In transitional areas where there is some air mass movement, ***cold*** or ***warm fronts*** can develop. **Figure 7r-2** illustrates a vertical cross-section of a ***cold front***. A cold front is the transition zone in the ***atmosphere*** where an advancing cold, dry stable ***air mass*** displaces a warm, moist unstable subtropical air mass. On a weather map, the cold front is drawn as a solid blue line with triangles. The position of the triangles shows the direction of frontal movement. Cold fronts move between 15 to 50 kilometers per hour in a southeast to east direction. The formation of clouds and precipitation at the frontal zone is caused by frontal lifting. High altitude ***cirrus*** clouds are found well in advance of the front. Above the surface location of the cold front, high altitude ***cirrostratus*** and middle altitude ***altocumulus*** are common. ***Precipitation*** is normally found just behind the front where frontal lifting has caused the development of towering ***cumulus*** and ***cumulonimbus*** clouds. **Table 7r-1** describes some of the weather conditions associated with a cold front.

Table 7r-1: Weather conditions associated with a cold front.

{PRIVATE}**Weather Phenomenon**	**Prior to the Passing of the Front**	**Contact with the Front**	**After the Passing of the Front**
Temperature	Warm	Cooling suddenly	Cold and getting colder
Atmospheric Pressure	Decreasing steadily	Leveling off then increasing	Increasing steadily
Winds	South to southeast	Variable and gusty	West to northwest
Precipitation	Showers	Heavy rain or snow, hail sometimes	Showers then clearing
Clouds	Cirrus and cirrostratus changing later to cumulus and cumulonimbus	Cumulus and cumulonimbus	Cumulus

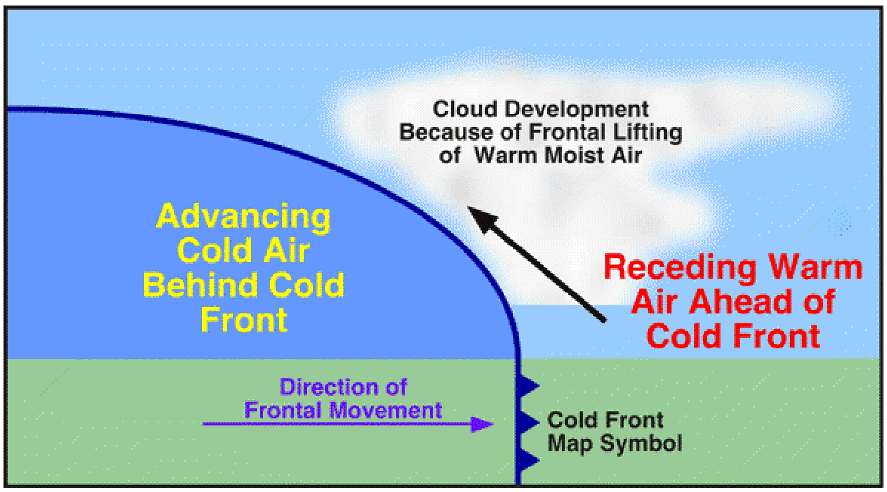


Figure 7r-2: Atmospheric cross-section of a cold front.

A ***warm front*** is illustrated in the cross-section **diagram** below (**Figure 7r-3**). A warm front is the transition zone in the ***atmosphere*** where an advancing warm subtropical, moist ***air mass*** replaces a retreating cold, dry polar air mass. On a weather map, a warm front is drawn as a solid red line with half-circles. The position of the half-circles shows the direction of frontal movement. Warm fronts move about 10 kilometers per hour in a northeast direction. This is less than half the speed of a cold front. The formation of clouds and precipitation ahead of the frontal zone is caused by gradual frontal lifting. High altitude ***cirrus***, ***cirrostratus*** and middle altitude ***altostratus*** clouds are found well in advance of the front. About 600 kilometers ahead of the front, ***nimbostratus*** clouds occur. These clouds produce ***precipitation*** in the form of snow or rain. Between the nimbostratus clouds and the surface location of the warm front, low altitude ***stratus*** clouds are found. Finally, a few hundred kilometers behind the front scattered ***stratocumulus*** are common in the lower ***troposphere***. **Table 7r-2** describes some of the weather conditions associated with a warm front.

Table 7r-2: Weather conditions associated with a warm front.

{PRIVATE}**Weather Phenomenon**	**Prior to the Passing of the Front**	**Contact with the Front**	**After the Passing of the Front**
Temperature	Cool	Warming suddenly	Warmer then leveling off
Atmospheric Pressure	Decreasing steadily	Leveling off	Slight rise followed by a decrease
Winds	South to southeast	Variable	South to southwest
Precipitation	Showers, snow, sleet or drizzle	Light drizzle	None
Clouds	Cirrus, cirrostratus, altostratus, nimbostratus, and then stratus	Stratus, sometimes cumulonimbus	Clearing with scattered stratus, sometimes scattered cumulonimbus

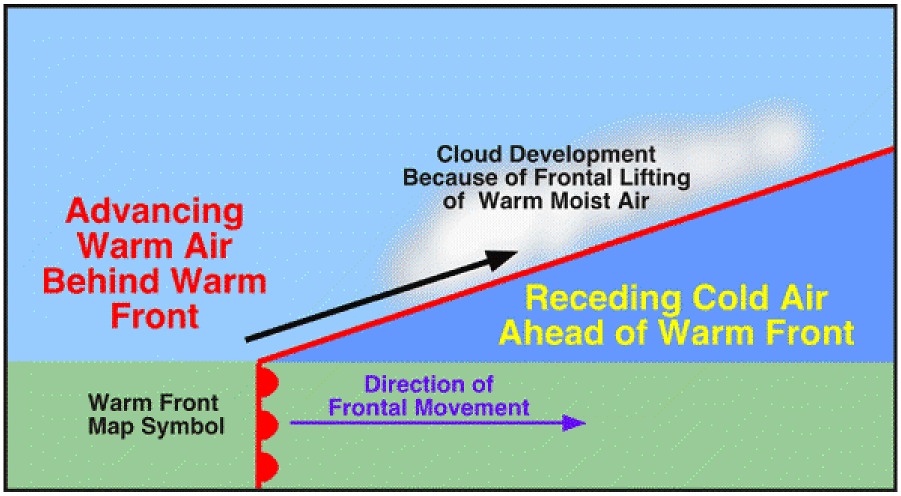


Figure 7r-3: Atmospheric cross-section of a warm front.

Occluded fronts are produced when a fast moving ***cold front*** catches and overtakes a slower moving ***warm front***. Two types of occluded fronts are generally recognized. A **cold type occluded front** occurs when the air behind the front is colder than the air ahead of the front. When the air behind the front is warmer than the air ahead of the front a **warm type occluded front** is produced. Warm type occlusions are common on the west coast of continents and generally form when maritime polar air collides with continental polar or arctic air. The cross-section **diagram** in **Figure 7r-4** illustrates a **cold type occlusion**. Note that in the occlusion process the invading mild moist air that was found behind the warm front has been lifted into the upper troposphere.

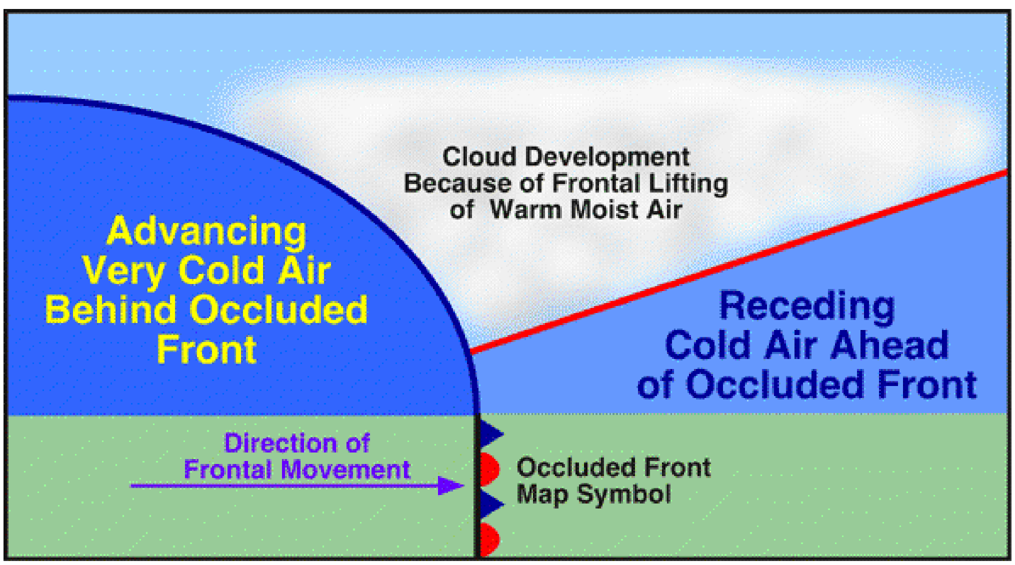


Figure 7r-4: Atmospheric cross-section of a occluded front.

Finally, the frontal systems described on this page are often associated with a storm system known as the ***mid-latitude cyclone***. The next section will describe this weather phenomenon in detail.

(s) The Mid-Latitude Cyclone

Introduction

Mid-latitude or ***frontal cyclones*** are large traveling atmospheric cyclonic storms up to 2000 kilometers in diameter with centers of low atmospheric pressure. An intense mid-latitude cyclone may have a surface pressure as low as 970 millibars, compared to an average sea-level pressure of 1013 millibars. Frontal cyclones are the dominant weather event of the Earth's mid-latitudes forming along the ***polar front***. (**Figure 7s-1**).

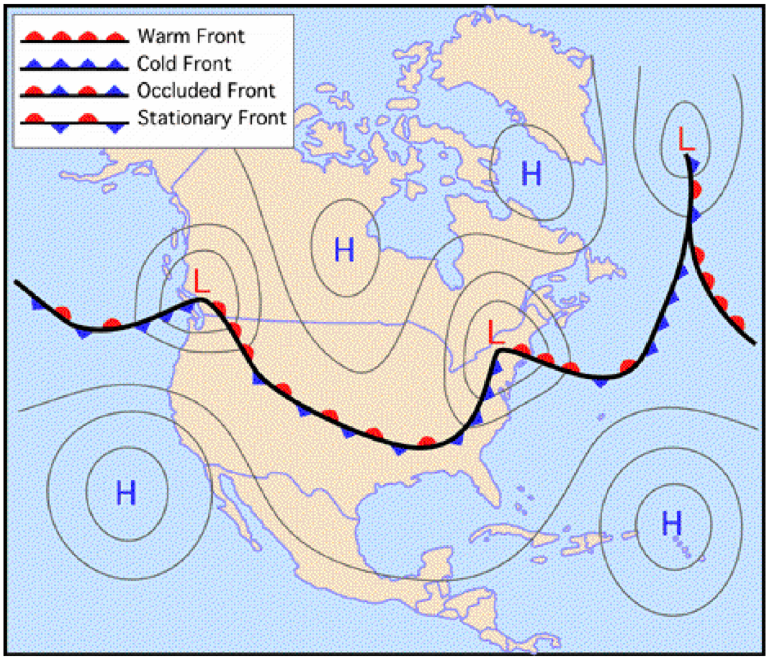
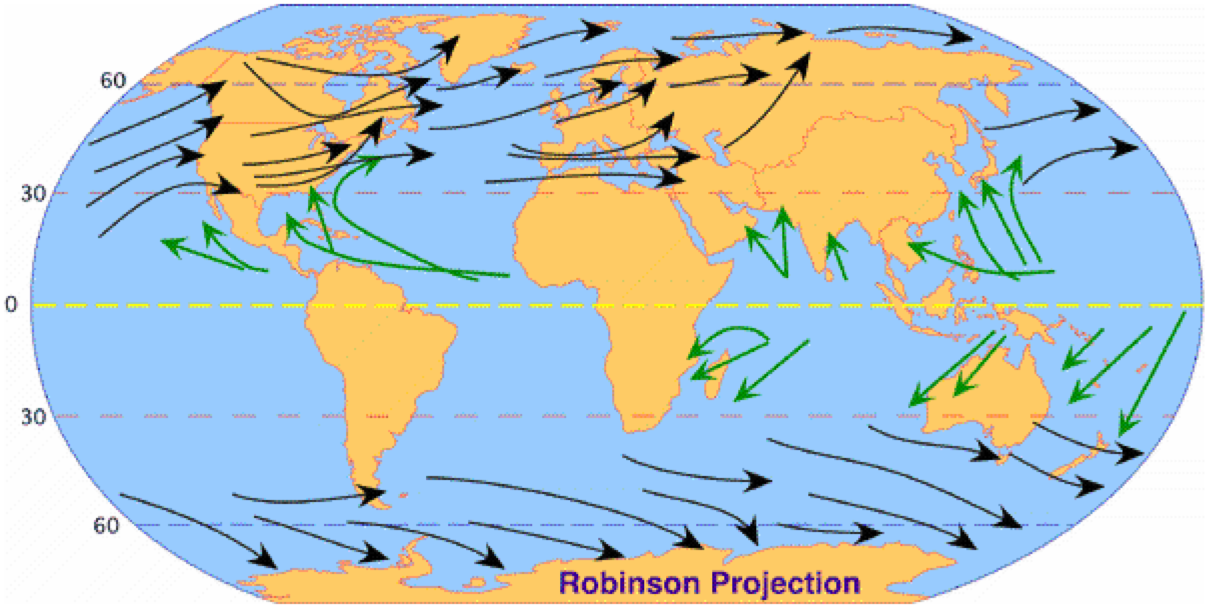


Figure 7s-1: A series of mid-latitude cyclones forming along the ***polar front*** (black line with red half circle and blue triangle symbols). On the illustration, the low pressure center of the mid-latitude cyclones is identfied by a red **L**. The systems located along the west and east coast of North America are in the middle stage of their life. The mid-latitude cyclone east of Greenland is at the end of its life cycle. In their mature stage, mid-latitude cyclones have a warm front on the east side of the storm's center and a cold front to the west. The cold front travels faster than the warm front. Near the end of the storm's life the cold front catches up to the warm front causing a condition known as ***occlusion***.

Mid-latitude cyclones are the result of the dynamic interaction of warm tropical and cold polar ***air masses*** at the ***polar front***. This interaction causes the warm air to be cyclonically lifted vertically into the atmosphere where it combines with colder upper atmosphere air. This process also helps to transport excess ***energy*** from the lower latitudes to the higher latitudes.

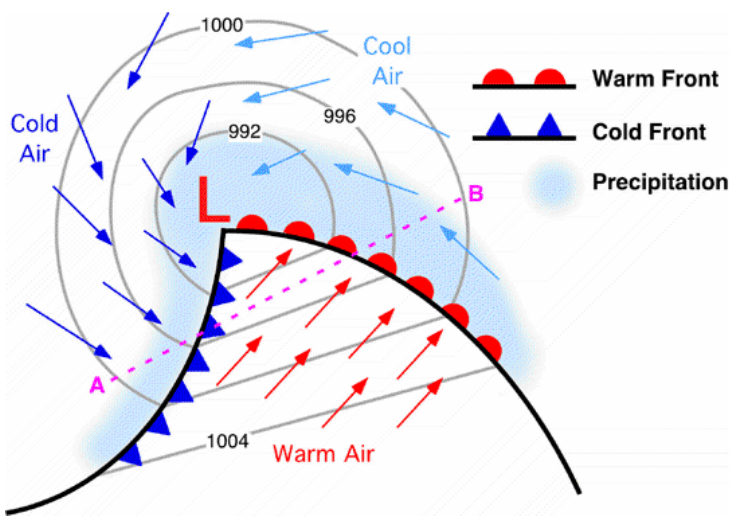
The mid-latitude cyclone is rarely motionless and commonly travels about 1200 kilometers in one day. Its direction of movement is generally eastward (**Figure 7s-2**). Precise movement of this weather system is controlled by the orientation of the ***polar jet stream*** in the upper troposphere. An estimate of future movement of the mid-latitude cyclone can be determined by the winds directly behind the cold front. If the winds are 70 kilometers per hour, the cyclone can be projected to continue its movement along the ground surface at this velocity.



{PRIVATE}Figure 7s-2: Typical paths of mid-latitude cyclones are represented by black arrows. This image also shows the typical paths traveled by subtropical ***hurricanes*** (green arrows).

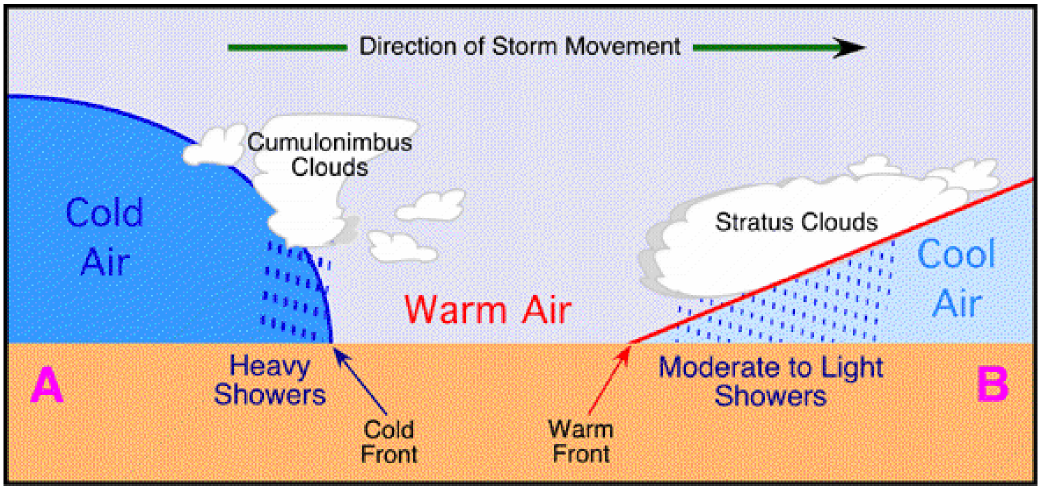
Figure 7s-3 describes the patterns of wind flow, surface pressure, fronts, and zones of precipitation associated with a mid-latitude cyclone in the Northern Hemisphere. Around the low, winds blow counterclockwise and inwards (clockwise and inward in the Southern Hemisphere). West of the low, cold air traveling from the north and northwest creates a ***cold front*** extending from the cyclone's center to the southwest. Southeast of the low, northward moving warm air from the subtropics produces a ***warm front***. ***Precipitation*** is located at the center of the low and along the fronts where air is being uplifted.

Mid-latitude cyclones can produce a wide variety of precipitation types. Precipitation types include: ***rain***, ***freezing rain***, ***hail***, ***sleet***, ***snow pellets***, and ***snow***. Frozen forms of precipitation (except hail) are common with storms that occur in the winter months. Hail is associated with severe thunderstorms that form along or in front of cold fronts during spring and summer months.



{PRIVATE}Figure 7s-3: Fronts, winds patterns, pressure patterns, and precipitation distribution found in an idealized mature mid-latitude cyclone.

Figure 7s-4 describes a vertical cross-section through a mature mid-latitude cyclone. In this cross-section, we can see how air temperature changes as we move from behind the ***cold front*** to a position ahead of the ***warm front***. Behind the surface position of the cold front, forward moving cold dense air causes the uplift of the warm lighter air in advance of the front. Because this uplift is relatively rapid along a steep frontal gradient, the condensed water vapor quickly organizes itself into ***cumulus*** and then ***cumulonimbus*** clouds. Cumulonimbus clouds produce heavy precipitation and can develop into **severe thunderstorms** if conditions are right (see **section *7t***). Along the gently sloping warm front, the lifting of moist air produces first ***nimbostratus*** clouds followed by ***altostratus*** and ***cirrostratus***. Precipitation is less intense along this front, varying from moderate to light showers some distance ahead of the surface location of the warm front.



{PRIVATE**}Figure 7s-4:** Vertical cross-section of the line **A-B** in **Figure 7s-3**.

Frontal cyclone development is related to polar jet stream processes. Within the jet stream, localized areas of air outflow can occur because of upper air divergence. Outflow results in the development of an upper air vacuum. To compensate for the vacuum in the upper atmosphere, surface air flows cyclonically upward into the outflow to replenish lost mass. The process stops and the mid-latitude cyclone dissipates when the upper air vacuum is filled with surface air.

Mid-latitude cyclones cause far less damage than **tropical cyclones** or ***hurricanes***. Hurricanes involve much greater amounts of atmospheric energy exchange. As one goes away from the equator, the energy available to fuel a weather system decreases as the amount of solar radiation and heat declines. Mid-latitude cyclones can have winds as strong as what is associated with a weak hurricane. But, this is a rare occurrence. Frontal cyclones tend to be most disruptive to human activity during winter months. Winter storms can produce heavy snowfalls or freezing rain which slows down transportation, snaps powerlines, and kills vegetation. In January 1998, a winter storm in eastern North America resulted in more than 20 human deaths, billions of dollars of damage, the loss of electrical power in some areas for up to two weeks, and the destruction of many deciduous trees because of the weight of ice (**Figures 7s-5**, **7s-6**, and **7s-7**).

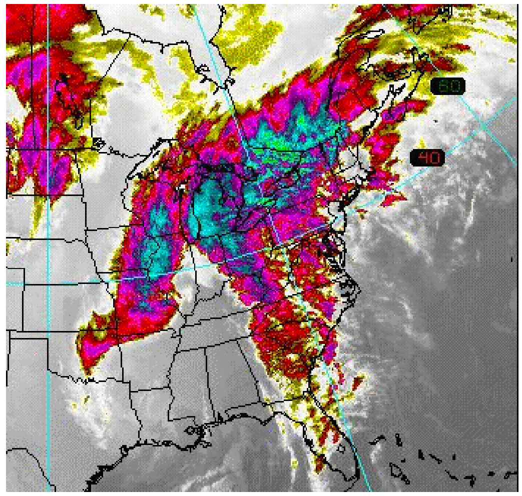


Figure 7s-5: ***GOES*** false color satellite image of the ice storm of January 1998. In the image, the center of the mid-latitude cyclone is located over the Great Lakes. This system pulled moisture from the Gulf of Mexico and the Atlantic Ocean which was converted into freezing rain and snow that fell from north-eastern United States to south-eastern Canada.

Figure 7s-6: The weight of ice collapsed many power-transmission towers leaving some areas without electricity for almost two weeks. (**Source**: Human Resources Development Canada - *Ice Storm '98 Emergency: A Study in Response*).

{PRIVATE**}Figure 7s-7:** Freezing rain from the 1998 ice storm also took its toll on many trees. (**Source**: Human Resources Development Canada - *Ice Storm '98 Emergency: A Study in Response*).

Cyclogenesis

The **animation** in **Figure 7s-8** illustrates the life cycle or ***cyclogenesis*** of the mid-latitude cyclone. The cyclone begins as a weak disturbance somewhere along the ***frontal zone*** (***stationary front***) where cold air from polar regions meets warm air from the south (**Stage 1**). The collision of these two air masses results in the uplift of the warm air into the upper atmosphere creating a cyclonic spin around a low pressure center (**Stages 2 and 3**). Associated with this center are the ***cold*** and ***warm fronts*** described in **topic *7r***. During the middle stages of cyclogenesis, the storm intensifies and the pressure at the storm's center drops (**Stages 4 and 5**). The warm air south of the low's center and between the two fronts is known as the **warm sector**. Cold fronts usually move along the Earth's surface at velocities greater than the warm front. As a result, the late stages of cyclogenesis occur when the cold front overtakes the warm front causing the air in the warm sector to be lifted into the upper atmosphere (**Stages 6 and 7**). The resulting boundary between the cold and cool air masses is called an ***occluded front***. A day or two after occlusion the occluded front dissipates, winds subside, and the stationary front forms on the surface of the Earth again (**Stages 8 to 10**).

(t) Thunderstor ms and Tornadoes

{PRIVATE}Thunderstorms

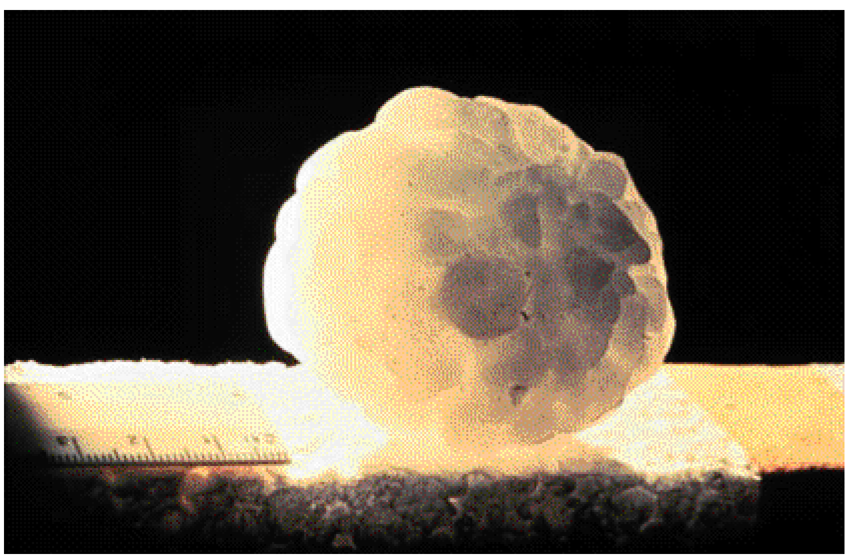
Thunderstorms form when moist, unstable air is lifted vertically into the atmosphere. Lifting of this air results in condensation and the release of ***latent heat***. The process to initiate vertical lifting can be caused by:

(1). Unequal warming of the surface of the Earth.

(2). ***Orographic lifting*** due to topographic obstruction of air flow.

(3). Dynamic lifting because of the presence of a ***frontal zone***.

Immediately after lifting begins, the rising parcel of warm moist air begins to cool because of ***adiabatic*** expansion. At a certain elevation the ***dew point*** is reached resulting in ***condensation*** and the formation of a ***cumulus*** cloud. For the cumulus cloud to form into a thunderstorm, continued uplift must occur in an unstable atmosphere. With the vertical extension of the air parcel, the cumulus cloud grows into a cumulonimbus cloud. ***Cumulonimbus*** clouds can reach heights of 20 kilometers above the Earth's surface. Severe weather associated with some these clouds includes ***hail***, strong winds, ***thunder***, ***lightning***, intense ***rain***, and ***tornadoes***.



{PRIVATE**}Figure 7t-1:** Hail stone measuring 21 centimeters in diameter. (**Source**: *NOAA Photo Collection Website*)**.**

{PRIVATE**}Figure 7t-2:** Multiple lightning strikes from a thunderstorm occurring at night. (**Source**: *NOAA Photo Collection Website*)**.**

Generally, two types of thunderstorms are common:

1) Air mass thunderstorms which occur in the mid-latitudes in summer and at the equator all year long.

2) Thunderstorms associated with ***mid-latitude cyclone cold fronts*** or ***dry lines***. This type of thunderstorm often has severe weather associated with it.

The most common type of thunderstorm is the air mass storm. Air mass thunderstorms normally develop in late afternoon hours when surface heating produces the maximum number of convection currents in the atmosphere. The life cycle of these weather events has three distinct stages. The first stage of air mass thunderstorm development is called the **cumulus stage (Figure 7t-3).** In this stage, parcels of warm humid air rise and cool to form clusters of puffy white ***cumulus*** clouds. The clouds are the result of condensation and deposition which releases large quantities of ***latent heat***. The added ***heat energy*** keeps the air inside the cloud warmer than the air around it. The cloud continues to develop as long as more humid air is added to it from below. Updrafts dominate the circulation patterns within the cloud.

{PRIVATE**}Figure 7t-3:** Developing thunderstorm cloud at the cumulus stage.

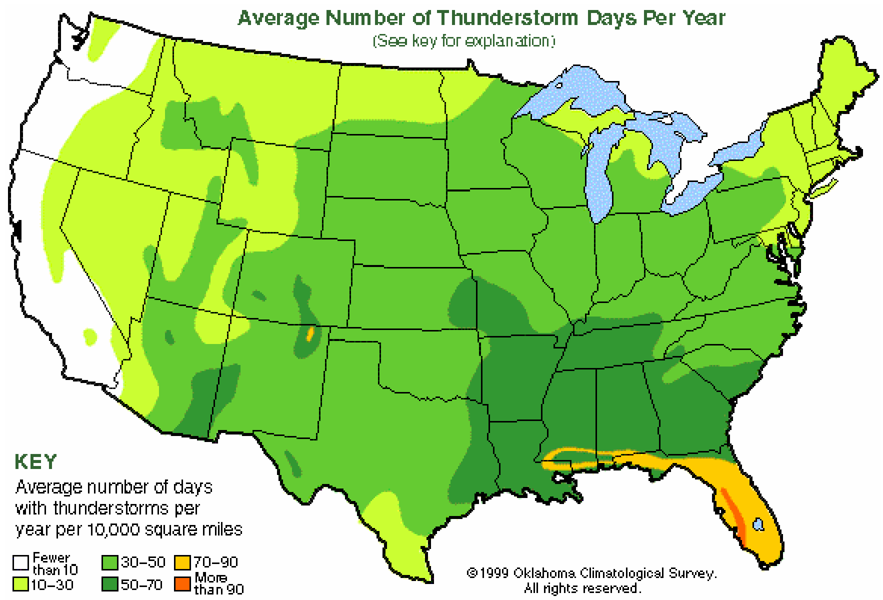
When the updrafts reach their maximum altitude in the developing cloud, usually 12 to 14 kilometers, they change their direction 180 degrees and become downdrafts. This marks the **mature stage**. With the downdrafts, precipitation begins to form through collision and coalescence. The storm is also at its most intense stage of development and is now a ***cumulonimbus*** cloud (**Figure 7t-4**). The top of the cloud takes on the familiar anvil shape, as strong stratospheric upper-level winds spread ice crystals in the top of the cloud horizontally. At its base, the thunderstorm is several kilometers in diameter. The mature air mass thunderstorm contains heavy rain, thunder, lightning, and produces wind gusts at the surface.

{PRIVATE**}Figure 7t-4:** Mature thunderstorm cloud with typical anvil shaped cloud. (**Source**: *NOAA Photo Collection Website*)**.**

The mature thunderstorm begins to decrease in intensity and enters the **dissipating stage** after about half an hour . Air currents within the convective storm are now mainly downdrafts as the supply of warm moist air from the lower atmosphere is depleted. Within about 1 hour, the storm is finished and precipitation has stopped.

Thunderstorms form from the equator to as far north as Alaska. They occur most commonly in the tropics were convectional heating of moist surface air occurs year round. Many tropical land based locations experience over 100 thunderstorm days per year. Thunderstorm formation over tropical oceans is less frequent because these surfaces do not warm rapidly. Outside the tropics, thunderstorm formation is more seasonal occurring in those months where heating is most intense.

Figure 7t-5 describes the annual average number of thunderstorm days across the United States. According to this map, the greatest incidence of thunderstorms occurs in the southeast and in parts of Colorado, Arizona, and New Mexico. This particular spatial distribution suggests that extreme solar heating is not the only requirement for thunderstorm formation. Another important prerequisite is the availability of warm moist air. In the United States, the Gulf of Mexico supplies adjacent continental areas with moist ***maritime tropical*** air masses. These air masses are relatively unstable quickly forming cumulonimbus clouds when surface heating is intense. The secondary maximums found in Colorado, Arizona, and New Mexico are due to another climatic factor. All of these areas are on the leeward side of the Rocky Mountains. Mountain slopes in these areas that face the sun absorb more direct solar radiation and become relatively warmer creating strong updrafts that form into cumulus clouds. If the differential heating is also supplemented by winds from the east, the cumulus clouds are further enhanced to become thunderstorms. Few thunderstorms occur along the west coast of the United States. This region is dominated by cool ***maritime polar*** air masses which suppress convectional uplift over land.



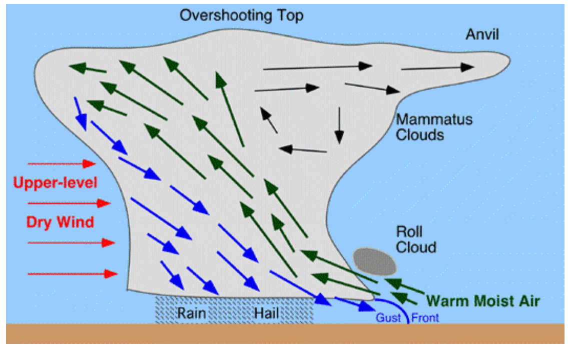
{PRIVATE}**Figure 7t-5:** Average number of thunderstorm days per year in the United States. (**Source:** *Oklahoma Climatological Survey*).

Severe Thunderstorms

Most thunderstorms are of the variety described above. However, some can form into more severe storms if the conditions exist to enhance and prolong the mature stage of development. Severe thunderstorms are defined as convective storms with frequent lighting, accompanied by local wind gusts of 97 kilometers per hour, or hail that is 2 centimeters in diameter or larger. Severe thunderstorms can also have tornadoes!

In most severe thunderstorms, the movement of the storm, in roughly an easterly direction, can refresh the storm's supply of warm humid air. With a continual supply of latent heat energy, the updrafts and downdrafts within the storm become balanced and the storm maintains itself indefinitely. Movement of the severe storm is usually caused by the presence of a ***mid-latitude cyclone cold front*** or a ***dry line*** some 100 to 300 kilometers ahead of a cold front. In the spring and early summer, frontal cyclones are common weather events that move from west to east in the mid-latitudes. At the same time, the ground surface in the mid-latitudes is receiving elevated levels of insolation which creates ideal conditions for air mass thunderstorm formation. When the cold front or dry line of a frontal cyclone comes in contact with this warm air it pushes it like a bulldozer both horizontally and vertically. If this air has a high humidity and extends some distance to the east, the movement of the mid-latitude cyclone enhances vertical uplift in storm and keeps the thunderstorms supplied with moisture and energy. Thus, the mid-latitude cyclone converts air mass thunderstorms into severe thunderstorms that last for many hours. Severe thunderstorms dissipate only when no more warm moist air is encountered. This condition occurs several hours after nightfall when the atmosphere begins to cool off.

Figure 7t-6 illustrates the features associated with a severe thunderstorm. This storm would be moving from left to right because of the motion associated with a mid-latitude cyclone. The upper-level dry air wind is generated from the mid-latitude cyclone. It causes the tilting of vertical air currents within the storm so that the updrafts move up and over the downdrafts. The **green arrows** represent the updrafts which are created as warm moist air is forced into the front of the storm. At the back end of the cloud, the updrafts swing around and become downdrafts (**blue arrows**). The leading edge of the downdrafts produces a ***gust front*** near the surface. As the gust front passes, the wind on the surface shifts and becomes strong with gusts exceeding 100 kilometers per hour, temperatures become cold, and the surface pressure rises. Warm moist air that rises over the gust front may form a ***roll cloud***. These clouds are especially prevalent when an inversion exists near the base of the thunderstorm.



{PRIVATE**}Figure 7t-6:** Model of the major features and circulation patterns associated with a severe thunderstorm.

Some severe thunderstorms develop a strong vertical updraft, commonly known as a ***mesocyclone***. Mesocyclones measure about 3 to 10 kilometers across and extend from the storm's base to its top. They are also found in the southwest quadrant of the storm. In some cases, mesocyclones can overshoot the top of the storm and form a cloud dome (**Figure 7t-6**). About half of all mesocyclones spawn tornadoes. When a tornado occurs, the mesocyclone lengthens vertically, constricts, and spirals down to the ground surface. Scientists speculate that mesocyclones form when strong horizontal upper air winds interact with normally occurring updrafts. The shearing effect of this interaction forces the horizontal wind to flow upward intensifying the updraft.

Tornadoes

A ***tornado*** is a vortex of rapidly moving air associated with some severe thunderstorms (see **Figure 7t-7**). Tornadoes that travel across lakes or oceans are called ***waterspouts***. Winds within the tornado funnel may exceed 500 kilometers per hour. High velocity winds cause most of the damage associated with these weather events. Tornadoes also cause damage through air pressure reductions. The air pressure at the tornado center is approximately 800 millibars (average sea-level pressure is 1014 millibars) and many human made structures collapse outward when subject to pressure drops of this magnitude. The destructive path of a tornado is usually about half a kilometer wide, and usually no more than 25 kilometers long. However, a spring tornado in 1917 traveled 570 kilometers across Illinois and Indiana lasting well over 7 hours.

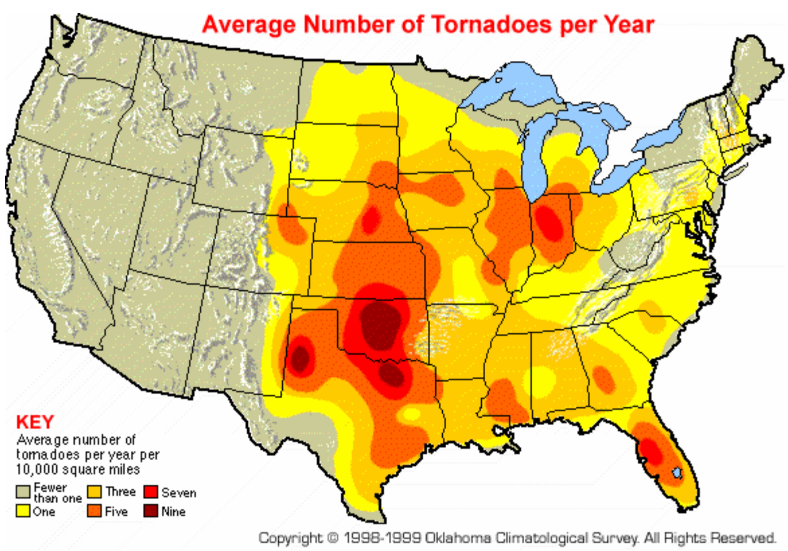
Figure 7t-7: Tornado.

About 74 % of all tornadoes have wind speeds between 65 and 181 kilometers per hour. These events are classified according to the ***Fujita tornado intensity scale*** as being weak (**Figure 7t-1**). Damage from these tornadoes varies from broken windows and tree branches to shingles blowing off roofs and moving cars pushed from roads. Weak tornadoes have a path that is about 1.5 kilometers long and 100 meters wide, and they generally last for only 1 to 3 minutes. According to the Fujita scale, strong tornadoes can have wind speeds between 182 and 332 kilometers per hour. These phenomena cause considerable damage and occur about 25 % of the time. Strong tornadoes can have a course up to 100 kilometers long and half a kilometer wide, and they can last for more than 2 hours. The rarest tornadoes are those with either a F4 or F5 rating. These events have wind speeds between 333 to 513 kilometers per hour and are very destructive and violent. F4 tornadoes occur only about 1 % of the time, while F5 are even more rare with a chance of about 1 in 1000 of happening.

Table 7t-1: Fujita tornado intensity scale.

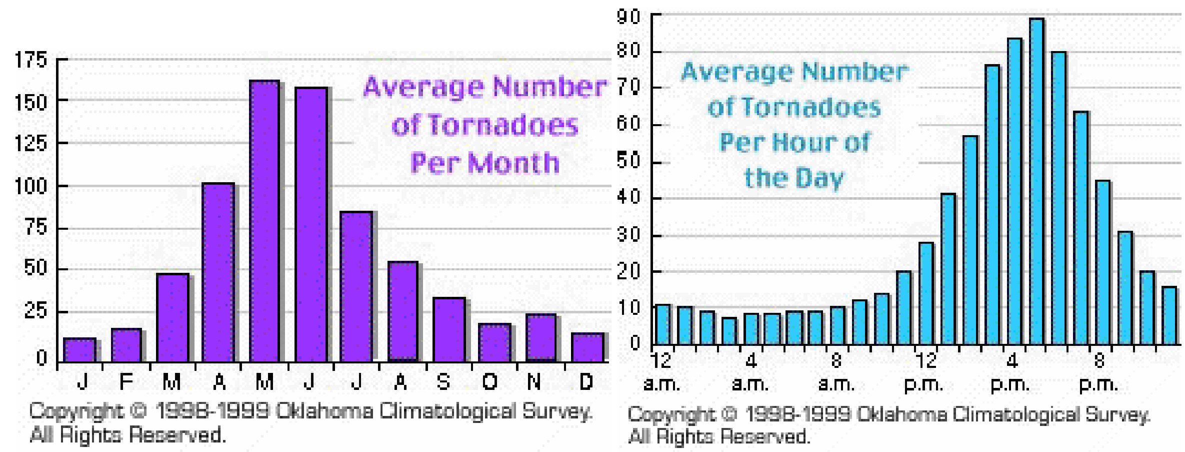
{PRIVATE}**F-Scale**	**Category**	**Kilometers per Hour (**Miles per Hour**)**	**Comments**
0	Weak	65-118 (40-73)	Damage is light. Chimneys on houses may be damaged; trees have broken branches; shallow-rooted trees pushed over; some windows broken; damage to sign boards.
1	Weak	119-181 (74-112)	Shingles on roofs blown off; mobile homes pushed off foundations or overturned; moving cars pushed off roads.
2	Strong	182-253 (113-157)	Considerable damage. Roofs torn off houses; mobile homes destroyed; train boxcars pushed over; large trees snapped or uprooted; light-objects thrown like missiles.
3	Strong	254-332 (158-206)	Damage is severe. Roofs and walls torn off better constructed homes, businesses, and schools; trains overturned; most trees uprooted; heavy cars lifted off ground and thrown some distance.
4	Violent	333-419 (207-260)	Better constructed homes completely leveled; structures with weak foundation blown off some distance.
5	Violent	420-513 (261-318)	Better constructed homes lifted off foundations and carried considerable distance where they disintegrate; trees debarked; cars thrown in excess of 100 meters.

Tornadoes occur in many parts of the world. Some notable hots spots include South Africa, Australia, Europe, New Zealand, northern India, Canada, Argentina, Uruguay, and the United States. Of these locations, the United States has some specific regions within its boundaries that have an extremely high number of events per year. **Figure 7t-8** describes the spatial distribution of tornado frequency in the United States. The highest number of tornadoes occurs in the southern plains (also called ***Tornado Alley***), south of Lake Michigan, and west central Florida.



{PRIVATE}**Figure 7t-8:** Average number of tornadoes per year in the United States. (**Source:** *Oklahoma Climatological Survey*).

Tornado occurrence has some interesting temporal characteristics. In the United States, most tornadoes occur in April, May, June, and July (**Figure 7t-9**). It is during these months that we get the conditions necessary for the formation of severe thunderstorms (see severe thunderstorm discussion above). Tornadoes also have particular times of the day in which they form. Most tornadoes form in the afternoon. Few develop in the late evening or early morning. During these times, heating of the Earth's surface is minimal or not occurring. Intense heating enhances updrafts in severe thunderstorms. It is these updrafts that can cause a tornado to form.



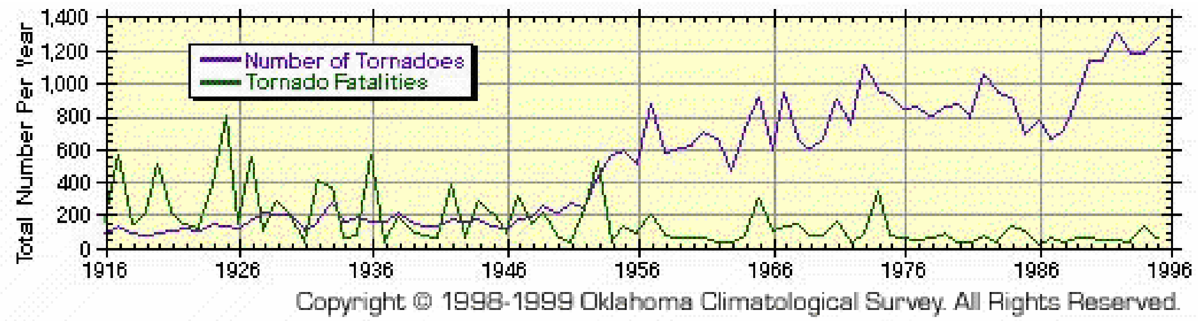
{PRIVATE**}Figure 7t-9:** Average number of tornadoes per month and per hour of the day in the United States. (**Source:** *Oklahoma Climatological Survey*).

In the United States, about 40,000 tornadoes have occurred in the last fifty years (1950-1999). Data for the period 1916 to 1996 indicates that the frequency of tornadoes in the United States has increased substantially (**Figure 7t-10**). However, part of this trend may be due to increased population densities. More people per unit area means a greater chance of seeing this relatively rare weather event. Two other factors that could also be responsible for the perceived increase in tornado numbers may be satellite imaging and weather radar. Both of these technologies allow us to pinpoint the thunderstorms that may generate funnels.

Total damage from tornadoes over the last 50 years has been estimated to be about 25 billion dollars. Tornadoes also take a heavy toll on human lives. **Table 7t-2** describes the ten deadliest tornado events in the United States. However, weather forecasting technology has played an important role in reducing the number of lives lost (**Figure 7t-10**). In the decade of the 1930s, before the advent of severe weather forecasting, 1945 people were killed by tornadoes. From 1986 to 1995, only 418 individuals perished suggesting a 90% decrease in fatalities. This reduction is even more astonishing when you consider that the population of the United States doubled from 1935 to 1990.

Table 7t-2: Ten deadliest tornado events in the United States.

{PRIVATE**}DATE**	**LOCATION(S)**	**DEATHS**
March 18, 1925	Missouri, Illinois, Indiana	689
May 6, 1840	Natchez, Mississippi	317
May 27, 1896	St. Louis, Missouri	255
April 5, 1936	Tupelo, Mississippi	216
April 6, 1936	Gainesville, Georgia	203
April 9, 1947	Woodward, Oklahoma	181
April 24, 1908	Amite, Louisiana and Purvis, Mississippi	143
June 12, 1899	New Richmond, Wisconsin	117
June 8, 1953	Flint, Missouri	115
May 11, 1953	Waco, Texas	114



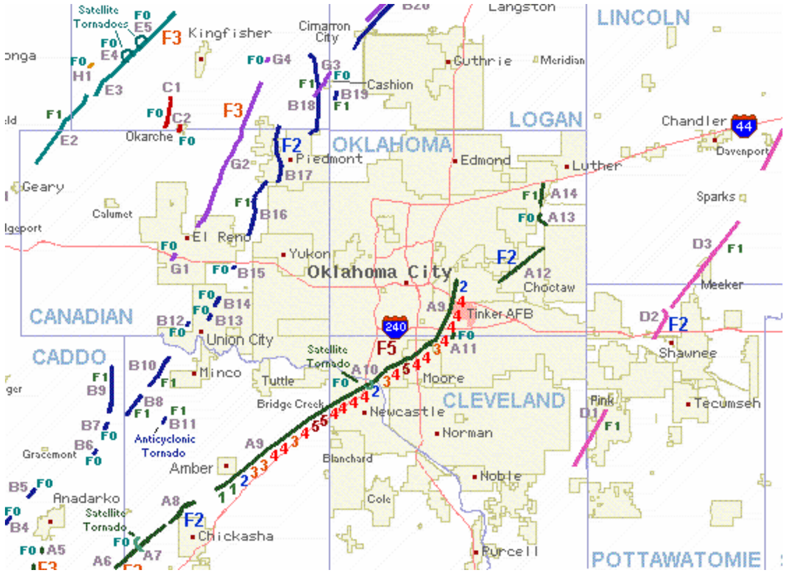
{PRIVATE**}Figure 7t-10:** Average number of tornadoes and tornado fatalities per year in the United States for the period 1916 to 1996. (**Source:** *Oklahoma Climatological Survey*).

Oklahoma, May 3, 1999

Tornadoes can cause considerable damage to natural and human made structures on the Earth's surface. Often, they also cause the injury and death of people. One particular event that set new monetary records in destruction occurred in Oklahoma on May 3, 1999. The destruction from this meteorological event began at 4:45 in the afternoon and ended about six hour later. During this time period, about 70 tornadoes were spawned along a 240-kilometer long swath that began in southwest Oklahoma.

Oklahoma receives more tornadoes per square kilometer than any other region found on our planet. It also receives most of these severe events in the month of May when warm moist air from the Gulf of Mexico interacts with ***cold fronts*** descending from the northwest. On the morning of May 3rd, forecasters at NOAA's ***Storm Prediction Center*** predicted only a slight risk for severe weather across parts of Oklahoma, Texas, and Kansas. However, by the afternoon conditions for tornado generation had improved greatly. Clearing skies provided the solar radiation required to increase surface temperatures and enhance convective development of severe thunderstorms. A kink in the jet stream, known as a ***short wave***, strengthened upper air winds traveling from the east. Strong upper air winds have a tendency to enhance the vertical ***mesocyclonic*** circulation occurring inside developing thunderstorms. This information was added to the high-resolution computer forecast model at the Storm Prediction Center. At 3:49, the Storm Prediction Center updated the forecast and issued a ***tornado watch***. Less than an hour later the damage began.

The greatest damage on May 3 occurred when the funnels cut through Oklahoma City (**Figure 7t-11**). The damage to the urban area stretched over 60 kilometers and was at times more than a kilometer wide. In some places, winds from the funnel were traveling about 420 to 510 kilometers per hour! More that 2500 buildings were destroyed and another 7500 were damaged. Total damage to structures and other property was estimated to be about 1.2 billion dollars. The toll on human life, for the entire state, was 40 individuals. This number would have certainly been much higher if radio and television warning systems did not exist.



{PRIVATE**}Figure 7t-11:** Tornado paths in and around Oklahoma City on May 3, 1999. Individual tornado tracks are identify by different colors. Tornado intensity is given by Fujita Scale values. Note that the largest tornado track, that cuts through the center of the city, has intensity values recorded along its entire path. (**Source:** NOAA *National Weather Services South Region - The Central Oklahoma Tornado Outbreak of May 3, 1999*).

(u) Tropical Weather and Hurricanes

{PRIVATE}Tropical Weather

The tropics can be defined as the area of the Earth found between the Tropic of Cancer and the Tropic Capricorn. In this region, the sun will be directly overhead during some part of the year. The temperature of the tropics does not vary much from season to season because considerable solar insolation is received by locations regardless of the season. Weather in the tropics is dominated by convective storms that develop mainly along the ***intertropical convergence zone***, the ***subtropical high pressure zone***, and oceanic disturbances in the ***trade winds*** that sometimes develop into ***hurricanes***.

One of the most important weather features found in the tropics is the ***intertropical convergence zone***. The intertropical convergence zone is distinguished by a wide band of ***cumulus*** and ***cumulonimbus*** clouds that are created by dynamic atmospheric lifting due to ***convergence*** and ***convection***.

Normally, the intertropical convergence zone delineates the location where the noonday sun is directly overhead on the globe. Because of the high sun, the intertropical convergence zone receives the greatest quantity of daily solar insolation in the tropics. At the intertropical convergence zone, this energy is used to evaporate large amounts of water and is coverted into sensible heat at the ground surface and within the atmosphere. Often, these processes lead to an almost daily development of convective ***thunderstorms*** by providing moisture and heat for the development of cumulonimbus clouds.

The intertropical convergence zone also represents the location of convergence of the northeast and southeast trade winds. The convergence of these wind systems enhances the development of convective rain clouds at the tropics.

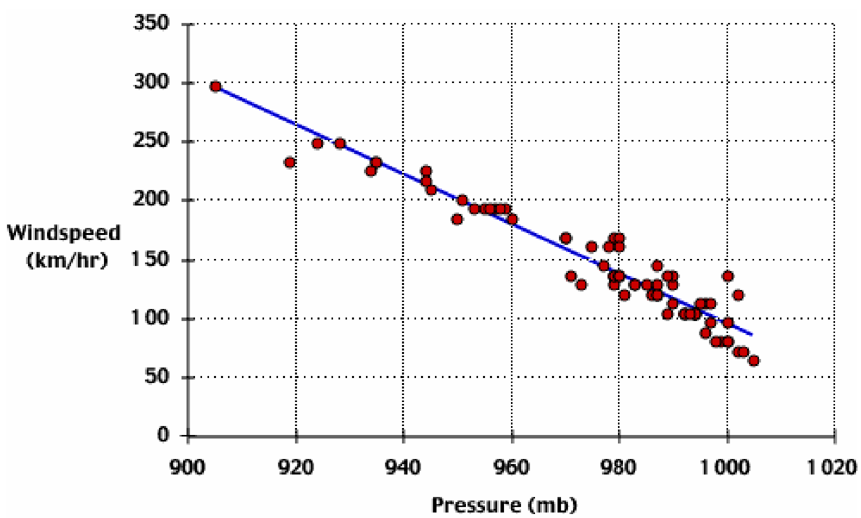
The intertropical convergence zone moves seasonally with the tilt of the Earth's axis. The convective rains that accompany the passage of the intertropical convergence zone are the primary source of precipitation for locations roughly 10 to 23.5 degrees North and South latutide. The animation below demonstrates the seasonal movement of this weather phenomenon.

Weather disturbances in the ***trade winds***, known as ***easterly waves***, are another source of cloud development and precipitation in the tropics. Easterly waves develop first as a weak ***disturbance*** in the atmosphere usually because of the presence of localized warmer sea temperatures. This is seen on a weather map as a wave in the ***isobars***. On the eastern side of the wave, ***convergence*** occurs forming thunderstorms (***divergence*** on the western side). If the convergence is strong enough the storm system may evolve into a hurricane.

The other important weather feature in the tropics is the ***subtropical high pressure zone***. Air flow in the subtropical high pressure zone is primarily descending. This creates clear skies, low humidities and hot daytime temperatures. Like the intertropical convergence zone, the subtropical high pressure zone migrates seasonally (see **Figure 7u-1** above). It generally influences locations 10 to 23.5 degrees North and South during some part of the year.

Hurricanes

Hurricanes are intense ***cyclonic*** storms that develop over the warm oceans of the tropics (**Figure 7u-3**). These tropical storms go by other names in the various parts of the world: India/Australia - cyclones; western North Pacific - typhoons; and the Philippines - baguio. By international agreement, the term **tropical cyclone** is used by most nations to describe hurricane-like storms that originated over tropical oceans. Surface atmospheric pressure in the center of a hurricane tends to be extreme low. The lowest pressure reading ever recorded for a hurricane (typhoon Tip, 1979) is 870 millibars (mb). However, most storms have an average pressure of 950 millibars. To be classified as a hurricane, sustained wind speeds must be greater than 118 kilometers per hour at the storm's center. Wind speed in a hurricane is directly related to the surface pressure of the storm. The following **graph** (**Figure 7u-2**) shows the relationship between surface pressure and sustained wind speed for a number of tropical low pressure systems.



{PRIVATE**}Figure 7u-2:** Relationship between surface pressure and wind speed for a number of tropical lowpressure systems. Tropical low pressure systems are classified as hurricanes when their pressure is 980 millibars or lower, and sustained wind speeds are greater than 118 kilometers per hour.

Hurricanes have no fronts associated with them like the ***mid-latitude cyclones*** of the ***polar front***. They are also smaller than the mid-latitude cyclone, measuring on average 550 kilometers in diameter. Mature hurricanes usually develop a cloud-free ***eye*** at their center (**Figure 7u-4**). In the eye, air is descending creating clear skies. The eye of the hurricane may be 20 to 50 kilometers in diameter. Surrounding the eye are bands of organized thunderstorm clouds formed as warm air move in and up into the storm. The strongest winds and heaviest precipitation are found in the area next to the eye where a vertical wall of thunderstorm clouds develops from the Earth's surface to the top of the ***troposphere***.

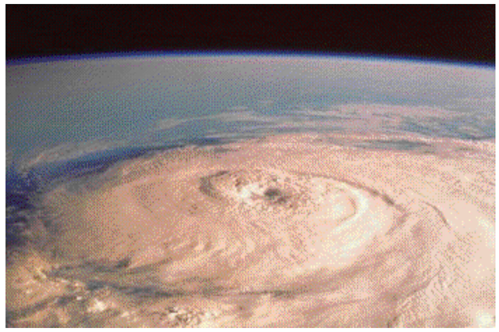
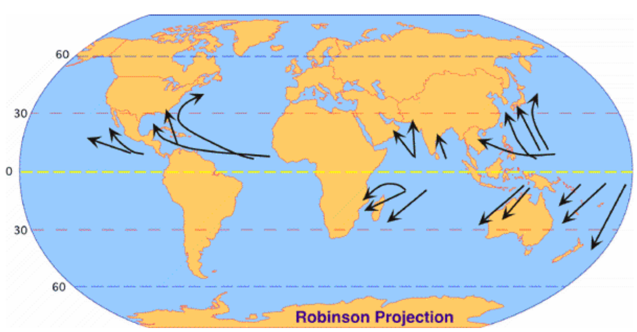


Figure 7u-3: Hurricane as seen from the space shuttle (**Source**: *NASA*).

{PRIVATE**}Figure 7u-4:** Satellite view of hurricane Floyd just before it made landfall in North Carolina, September 15, 1999. Notice the eye is clearly visible in this image

Hurricanes are powered by the ***latent heat energy*** released from ***condensation***. To form and develop they must be supplied with a constant supply of warm humid air for this process. Surface air with enough energy to generate a hurricane only exists over oceans with a temperature greater than 26.5 degrees Celsius. Ocean temperatures this high only occur in selected regions and during particular seasons. Hurricane development does not occur if a ***temperature inversion*** exists in the atmosphere. Inversions develop in the tropics when ***subtropical high pressure systems*** produce sinking air. Also, hurricanes do not develop in the region 5 degrees either side of the equator. Within this region ***Coriolis force*** is negligible. Coriolis force is required for the initiation of cyclonic flow.

Hurricanes dissipate when their energy supply is substantially reduced. This occurs either with landfall or storm movement into cooler seas. Most hurricanes live for about a week. However, if a hurricane remains over warm water its life can be extended. In 1992, hurricane Tina was an active tropical storm for 24 days over the North Pacific. The **map** in **Figure 7u-5** shows the typical areas of hurricane development and the usual paths they take during their life.



{PRIVATE**}Figure 7u-5:** Typical areas where hurricanes begin their development and the common paths of storm movement.

Hurricanes go through a number of different stages of development. Initially, these powerful storms begin their lives as an unorganized group of thunderstorms that develop over the specific areas of the tropical oceans. However, not all of these types of ***tropical disturbances*** become hurricanes. To develop into a hurricane, significant ***cyclonic*** circulation must occur around the disturbances. This type of circulation enhances the development of the group of thunderstorms by providing additional moisture and latent ***heat energy***. With more moisture and latent heat energy, the strength and number of the thunderstorms in the tropical disturbance increases causing the disturbance to intensify. The thunderstorms also begin to organize themselves into spiral bands that swirl cyclonically toward the center of the storm. If the sustained wind speed around the disturbance increases to between 37 and 63 kilometers per hour, the storm becomes classified as a ***tropical depression***. Tropical depressions appear on the weather map as a cyclonic low with several closed ***isobars*** circling the storm's center. A tropical depression can continue to intensify and become a ***tropical storm*** which is the next stage in hurricane development. Tropical storms have a lower central pressure, several more closed isobars on a weather map, and winds that are between 64 and 118 kilometers per hour. Finally, tropical storms officially become hurricanes when their sustained wind speed exceeds 118 kilometers per hour.

Tropical storms and hurricanes are the most deadly and destructive type of severe weather on our planet. One of the most destructive hurricanes this century was Andrew in August of 1992. This storm, which had a minimal pressure reading of 922 millibars, caused an estimated 26 billion dollars worth of damage mainly in Dade County, Florida. The deadliest storm this century is probably hurricane Mitch which hit Central America in late October and early November 1998. Estimates suggest that over 11,000 people died from this storm. Most of these deaths were cause by flooding and mudslides due to heavy rains.

The damage that hurricanes inflict is caused by high wind speeds, heavy rainfall, storm surge and tornadoes. Wind speed in a hurricane is usually directly related to atmospheric pressure (see **Figure 7u-2** above). The lower the pressure the faster the winds blow. Wind speed also varies within the storm. As discussed earlier, winds are usually strongest at the edge of the hurricane's eye. High winds inflict damage by blowing down objects, creating choppy waves and high seas, and by inundating coastal areas with seawater. Rainfall within a hurricane can often exceed 60 centimeters in a 24 hour period. If this rainfall occurs on land, flooding normally occurs. ***Storm surge*** is increase in the height of the ocean's surface in the region beneath and around the eye of the storm. It occurs when low atmospheric pressure causes the ocean surface to expand and because the hurricane's cyclonic winds blow seawater towards the eye. Hurricane Camille (1969) had a storm surge of more than seven meters with a central pressure of 909 millibars. A considerable amount of damage can also occur because of hurricane generated tornadoes. About 25 % of the hurricanes that make landfall have tornadoes. Some scientists also suspect that the thunderstorms that occur near the eye of a hurricane can produce very strong downbursts (vertical downward movements of air).

The year 1998 was extremely bad for the development of tropical storms and hurricanes in the Atlantic ocean. During the late summer and fall of that year fourteen tropical storms developed of which ten became hurricanes. **Table 7u-1** describes some of the characteristics of each of the storms that developed into hurricanes. Seven of the tropical storms and hurricanes made landfall in the United States causing about 6.5 billion dollars worth of damage. The strongest hurricane of the 1998 season was hurricane Mitch. This storm made landfall in Honduras and killed over 11,000 people in Central America.

Table 7u-1: Characteristics of the Atlantic hurricanes of the 1998 season.

{PRIVATE}**Name of Hurricane**	**Born**	**Died**	**Lifespan**	**Maximum Sustained Winds**	**Minimum Storm Pressure**	**Damage US Dollars**	**inDirect Deaths**
Bonnie (see ***movie***)	Aug 19	Aug 30	10.5 days	185 km/hr	954 mb	1 Billion	3
Danielle	Aug 24	Sept 3	10.5 days	169 km/hr	955 mb	Minimal (Bermuda)	None
Earl	Aug 31	Sept 3	3 days	161 km/hr	986 mb	80 Million	3
Georges	Sept 15	Sept 29	14 days	241 km/hr	938 mb	5.1 Billion	602
Ivan	Sept 20	Sept 26	6.5 days	145 km/hr	975 mb	Minimal (Azores)	None
Jeanne	Sept 21	Sept 30	9 days	169 km/hr	970 mb	Minimal (Azores)	None
Karl	Sept 23	Sept 27	4 days	169 km/hr	970 mb	None	None
Lisa	Oct 5	Oct 9	4 days	121 km/hr	987 mb	None	None
Mitch	Oct 21	Nov 5	14.5 days	290 km/hr	906 mb	Over 10 Billion	Over 11,000
Nicole	Nov 24	Dec 1	7 days	137 km/hr	979 mb	None	None

(v) Climate Classification and Climatic Regions of the World

{PRIVATE**}Climate Classification**

The **Köppen Climate Classification System** is the most widely used system for classifying the world's climates. Its categories are based on the annual and monthly averages of temperature and precipitation. The Köppen system recognizes five major climatic types; each type is designated by a capital letter.

A - Tropical Moist Climates: all months have average temperatures above 18 degrees Celsius.

B - Dry Climates: with deficient precipitation during most of the year.

C - Moist Mid-latitude Climates with Mild Winters.

D - Moist Mid-Latitude Climates with Cold Winters.

E - Polar Climates: with extremely cold winters and summers.

Tropical Moist Climates (A)

Tropical moist climates extend northward and southward from the equator to about 15 to 25 degrees of latitude. In these climates all months have average temperatures greater than 18 degrees Celsius. Annual ***precipitation*** is greater than 1500 mm. Three minor Köppen climate types exist in the A group, and their designation is based on seasonal distribution of rainfall. **Af** or **tropical wet** is a tropical climate where precipitation occurs all year long. Monthly temperature variations in this climate are less than 3 degrees Celsius. Because of intense surface heating and high humidity, ***cumulus*** and ***cumulonimbus*** clouds form early in the afternoons almost every day. Daily highs are about 32 degrees Celsius, while night time temperatures average 22 degrees Celsius. **Am** is a **tropical monsoon** climate. Annual rainfall is equal to or greater than **Af**, but most of the precipitation falls in the 7 to 9 hottest months. During the dry season very little rainfall occurs. The **tropical wet and dry** or savanna (**Aw**) has an extended dry season during winter. Precipitation during the wet season is usually less than 1000 millimeters, and only during the summer season.

Dry Climates (B)

The most obvious climatic feature of this climate is that ***potential evaporation*** and ***transpiration*** exceed ***precipitation***. These climates extend from 20 - 35 degrees North and South of the equator and in large continental regions of the mid-latitudes often surrounded by mountains. Minor types of this climate include:

Bw - dry arid (desert) is a true desert climate. It covers 12 % of the Earth's land surface and is dominated by xerophytic vegetation.

Bs - dry semiarid (steppe). Is a grassland climate that covers 14% of the Earth's land surface. It receives more precipitation than the **Bw** either from the ***intertropical convergence zone*** or from ***mid-latitude cyclones***.

Moist Subtropical Mid-Latitude Climates (C)

This climate generally has warm and humid summers with mild winters. Its extent is from 30 to 50 degrees of latitude mainly on the eastern and western borders of most continents. During the winter, the main weather feature is the ***mid-latitude cyclone***. Convective ***thunderstorms*** dominate summer months. Three minor types exist: **Cfa - humid subtropical**; **Cs - Mediterranean**; and **Cfb - marine**. The humid subtropical climate (**Cfa**) has hot muggy summers and frequent thunderstorms. Winters are mild and precipitation during this season comes from mid-latitude cyclones. A good example of a **Cfa** climate is the southeastern USA. **Cfb** marine climates are found on the western coasts of continents. They have a humid climate with short dry summer. Heavy precipitation occurs during the mild winters because of the continuous presence of mid-latitude cyclones. Mediterranean climates (**Cs**) receive rain primarily during winter season from the mid-latitude cyclone. Extreme summer aridity is caused by the sinking air of the ***subtropical highs*** and may exist for up to 5 months. Locations in North America are from Portland, Oregon to all of California.

Moist Continental Mid-latitude Climates (D)

Moist continental mid-latitude climates have warm to cool summers and cold winters. The location of these climates is poleward of the C climates. The average temperature of the warmest month is greater than 10 degrees Celsius, while the coldest month is less than -30 degrees Celsius. Winters are severe with snowstorms, strong winds, and bitter cold from Continental Polar or Arctic ***air masses***. Like the C climates there are three minor types: **Dw - dry winters**; **Ds - dry summers**; and **Df - wet all seasons**.

Polar Climates (E)

Polar climates have year-round cold temperatures with the warmest month less than 10 degrees Celsius. Polar climates are found on the northern coastal areas of North America, Europe, Asia, and on the landmasses of Greenland and Antarctica. Two minor climate types exist. **ET** or **polar tundra** is a climate where the soil is permanently frozen to depths of hundreds of meters, a condition known as permafrost. Vegetation is dominated by mosses, lichens, dwarf trees and scattered woody shrubs. **EF** or **polar ice caps** has a surface that is permanently covered with snow and ice.

Factors Influencing the World Climatic Regions

So far in this course we have discovered that the climate of a particular place is the function of a number of factors. These factors include:

1) Latitude and its influence on ***solar radiation*** received.
2) ***Air mass*** influences.
3) Location of global ***high*** and ***low pressure*** zones.
4) Heat exchange from ocean currents.
5) Distribution of mountain barriers.
6) Pattern of prevailing winds.
7) Distribution of land and sea.
8) Altitude.

At a macro-level, the first three factors are most important in influencing a region's climate. The animated **graphic** below provides us with a generalized model of the Earth's annual climatic variations. It also describes the latitudinal effects of these top three factors through the following climatic features:

Relative annual latitudinal location of the overhead sun at solar noon.

Intertropical convergence zone and its area of uplift, cloud development and precipitation.

Subtropical high pressure zone and its associated descending air currents and clear skies.

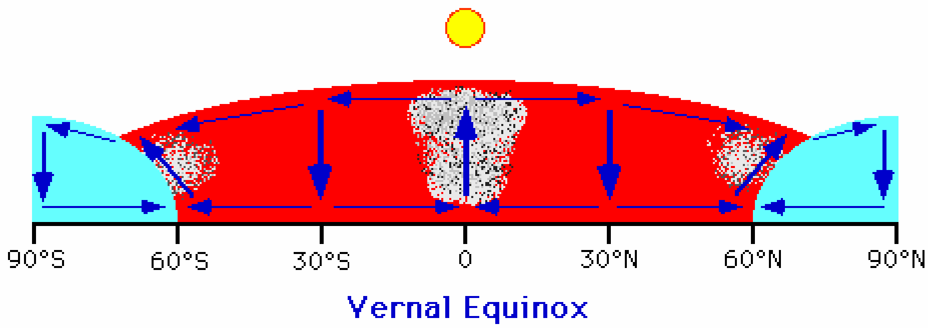
Polar front and its area of uplift, cloud development and precipitation.

Polar vortex and its associated descending air currents and clear skies.

Relative location of tropical/subtropical (**red**) and polar (**light blue**) air masses.

In the animation we can see that the intertropical convergence zone, the subtropical high pressure zone, polar front and the position of tropical/subtropical and polar air masses all move in response to the seasonal movements of the sun. It is important to understand this concept because of its climatic ramifications for locations on the globe. The type of climate that a location experiences is to a large extent a function of seasonal migration of these weather features. For example, a location at 15 degrees North latitude is influenced by the subtropical high pressure zone during winter solstice and by the intertropical convergence zone during the summer solstice. Another location, at 60 degrees North latitude, would be influenced by polar air masses during the winter solstice, the polar front during the equinoxes, and by subtropical air masses and the subtropical high pressure zone during the summer solstice.

When studying about the **Earth's Climatic Regions** as described below, use this **animation** (**Figure 7v-1**) as a guide to understanding the large scale climatic processes that produce each location's particular climate.



{PRIVATE}**Figure 7v-1:** Solar influence on the movement of the Earth's global circulation patterns.

Climatic Region Descriptions

The following discussion organizes the climatic regions of the world into eight different groups. Categorization of these climates is based on their **Köppen classification** and seasonal dominance of ***air masses***.

Tropical Wet

Köppen Classification - **Af**.

Dominated by Maritime Tropical air masses all year long.

The tropical wet climate is characterized by somewhat consistent daily high temperatures ranging between 20 to 30 degrees Celsius. The monthly temperature averages vary from 24 to 30 degrees Celsius. Annual range of monthly temperatures is about 3 degrees Celsius. It has reasonably uniform precipitation all year round, and total rainfall over 2000 millimeters or greater.

The region experiencing this climate lies within the effects of the ***intertropical convergence zone*** all year long. ***Convergence*** and high maritime humidities create ***cumulus clouds*** and ***thunderstorms*** almost daily.

Andagoya, Columbia 5 degrees N , Elevation: 65 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	27	27	28	28	27	27	27	27	27	27	27	27	27
Precip. mm	554	519	557	620	655	655	572	574	561	563	563	512	6905

Iquitos, Peru 4 degrees S , Elevation: 104 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	27	27	27	27	26	26	25	27	27	27	27	27	26
Precip. mm	256	276	349	306	271	199	165	157	191	214	244	217	2845

Tropical Wet and Dry

Köppen Classification - **Aw**, **Am** and **BS**.

Maritime Tropical air masses high sun season and Continental Tropical air masses low sun season.

This climate has distinct wet/dry periods. The seasonal pattern of moisture is due to the migration of the ***intertropical convergence zone***. The wet season is synchronous with the high sun and the presence of the convergence zone. The dry season is a result of the more stable air developing from the subsidence associated with the presence of the ***subtropical high zone*** during the low sun season.

During the rainy season, the climate of this location is similar to the tropical wet climate: warm, humid, and has frequent thunderstorms. During the dry season more or less semi-desert conditions prevail. Some regions may experience intensification of rainfall because of ***monsoon*** development and ***orographic uplift***.

Calcutta, India 22.5 degrees N , Elevation: 6 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	20	23	28	30	31	30	29	29	30	28	24	21	27
Precip. mm	13	24	27	43	121	259	301	306	290	160	35	3	1582

Mangalore, India 13 degrees N , Elevation: 22 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	27	27	28	29	29	27	26	26	26	27	27	27	27
Precip. mm	5	2	9	40	233	982	1059	577	267	206	71	18	3467

Cuiaba, Brazil 13.5 degrees S , Elevation: 165 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	27	27	27	27	26	24	24	26	28	28	28	27	27
Precip. mm	216	198	232	116	52	13	9	12	37	130	165	195	1375

Darwin, Australia 12.5 degrees S , Elevation: 27 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	28	28	28	28	27	25	25	26	28	29	29	29	28
Precip. mm	341	338	274	121	9	1	2	5	17	66	156	233	1563

Tropical Desert

Köppen Classification **- BW**. Dominated by Continental Tropical air masses all year.

This climate type covers 25 percent of all land area on the continents. The heart of the tropical desert climate is found near the tropics of Cancer and Capricorn, usually toward the western side of the continents. Regions with this climate have the following common climatic characteristics:

low relative humidity and cloud cover.

low frequency and amount of precipitation.

high mean annual temperature.

high monthly temperatures.

high diurnal temperature ranges.

high wind velocities.

The tropical desert climate is influenced by upper air stability and subsidence which is the result of the presence of the ***subtropical high pressure zone***. ***Relative humidity*** is normally low, averaging 10 to 30 percent in interior locations. Precipitation is very low in quantity and very infrequent in distribution, both temporally and spatially.

Temperature varies greatly both diurnally and annually. The highest average monthly temperatures on the Earth are found in the tropical desert. They range between 29 to 35 degrees Celsius. Winter monthly temperatures can be 15 to 25 degrees cooler than summer temperatures. This climate also has extreme diurnal ranges of temperature. The average diurnal range is from 14 to 25 degrees Celsius.

Wadi Halfa, Sudan 22 degrees N , Elevation: 160 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	15	17	21	26	31	32	32	33	30	28	22	17	25
Precip. mm	0	0	0	0	1	0	1	0	0	1	0	0	3

Berbera, Somalia 10.5 degrees N , Elevation: 8 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	25	26	27	29	32	37	37	37	34	29	26	26	30
Precip. mm	8	2	5	12	8	1	1	2	1	2	5	5	52

Alice Springs, Australia 23.5 degrees S , Elevation: 579 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	28	28	25	20	15	12	12	14	18	23	26	27	21
Precip. mm	44	34	28	10	15	13	7	8	7	18	29	39	252

Mid-Latitude Wet

Köppen Classification - **Cf** and **Df**.

Maritime Tropical in summer and Maritime Polar in winter.

The Mid-Latitude Wet climate is found in the Northern Hemisphere in the region from 60 degrees North to 25 to 30 degrees North mainly along the eastern margins of the continents. In North America, this climate extends from the Pacific coast of Canada at latitudes above 55 degrees eastward to the Atlantic coast where it dominates the eastern half of the continent. In the Southern Hemisphere, this climate exists on the Southeastern tip of South America, New Zealand and the Southeast coast of Australia.

Summer weather is dominated by Maritime Tropical ***air masses*** which produce many ***thunderstorms*** from daytime heating. Monthly average temperature ranges from 21 to 26 degrees Celsius with the tropical areas going as high as 29 degrees Celsius. This is slightly warmer than the humid tropics. Frontal weather associated with the ***mid-latitude cyclone*** dominates the climate of more polar areas and is more frequent in all regions in the winter.

Precipitation in this climate is fairly evenly distributed throughout the year. Annual totals of precipitation are quite variable and depend on the latitude and continental position of the regions. During the summer and on the equatorial margins, convectional rainfall is the primary mechanism of precipitation. The southeast of the United Sates averages 40 to 60 days of thunderstorms per year. The frequency of thunderstorms decreases rapidly from south to north. ***Hurricanes*** also provide a mechanism for producing precipitation in more tropical regions of this climate.

New Orleans, USA 30 degrees N , Elevation: 1 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	12	13	16	20	24	27	28	28	26	21	16	13	20
Precip. mm	98	101	136	116	111	113	171	136	128	72	85	104	1371

Williston, North Dakota 47.5 degrees N , Elevation: 579 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-12	-10	-4	6	13	17	22	20	14	8	-2	-8	5
Precip. mm	14	12	18	24	36	84	48	38	28	19	15	13	349

Buenos Aires, Argentina 34.5 degrees S , Elevation: 27 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	23	23	21	17	13	9	10	11	13	15	19	22	16
Precip. mm	103	82	122	90	79	68	61	68	80	100	90	83	1026

London, England 51.5 degrees N , Elevation: 5 m

{PRIVATE}	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	4	4	7	9	12	16	18	17	15	11	7	5	10
Precip. mm	54	40	37	38	46	46	56	59	50	57	64	48	595

Mid-Latitude Winter-Dry

Köppen Classification - **Cw** and **Dw**.

Maritime Tropical air masses in summer and Continental Polar air masses in winter.

This climate is characterized by a strong seasonal pattern of both temperature and precipitation. The normal location of the Mid-Latitude Winter-Dry climate is in the interior of the continents in the mid-latitudes. This continental location causes a large annual temperature range because of continentality.

This climate receives Maritime Tropical ***air masses*** in the summer with occasional Continental Tropical air masses from the adjacent deserts. Summers are hot and humid with intense summer convectional storms. Continental Polar air masses are dominant in the winter with an occasional outbreak of Maritime Polar air. Continental Polar air masses are associated with cold, dry weather conditions. Precipitation mainly occurs in the summer from ***thunderstorm*** activity. The mid-latitude cyclone produces a smaller quantity of precipitation in the winter.

Calgary, Canada 51 degrees N , Elevation: 1140 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-10	-9	-4	4	10	13	17	15	11	5	-2	-7	4
Precip. mm	17	20	26	35	52	88	58	59	35	23	16	15	444

Omaha, USA 41 degrees N , Elevation: 298 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-5	-3	3	11	17	23	26	25	19	13	4	-2	11
Precip. mm	21	24	37	65	88	115	86	101	67	44	32	20	700

Mid-Latitude Summer-Dry

Köppen Classification - **Cs** and **Ds**.

Summer weather is dominated by Continental Tropical air, while in the winter, Maritime Polar air masses are frequent.

The Mid-Latitude Summer-Dry climate is found on the western margins of the continents between 30 to 40 degrees of latitude. Usually, this climate does not spread into the continents very far. This climate is often called a Mediterranean climate.

Precipitation falls mainly in the winter in this climate via the ***mid-latitude cyclone***. During the summer these areas are influenced by stable ***subtropical highs***, that give them dry, warm weather.

Santiago, Chile 33.5 degrees S , Elevation: 512 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	19	19	17	13	11	8	8	9	11	13	16	19	14
Precip. mm	3	3	5	13	64	84	76	56	30	13	8	5	360

Los Angeles, USA 34 degrees N , Elevation: 37 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	13	14	15	17	18	20	23	23	22	18	17	15	18
Precip. mm	78	85	57	30	4	2	0	1	6	10	27	73	373

Rome, Italy 42 degrees N , Elevation: 131 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	8	8	10	13	17	22	24	24	21	16	12	9	15
Precip. mm	76	88	77	72	63	48	14	22	70	128	116	106	881

Polar Wet and Dry

Köppen Classification - **ET**.

Maritime Polar in summer and Continental Polar or Arctic in winter.

The polar wet and dry climate is characterized by cold winters, cool summers, and a summer rainfall regime. Areas experiencing this climate are the North American Arctic coast, Iceland, coastal Greenland, the Arctic coast of Europe and Asia, and the Southern Hemisphere islands of McQuarie, Kerguelen, and South Georgia. Annual precipitation averages less than 250 mm for most locations and most of this precipitation falls during the summer.

Isachsen, Canada 79 degrees N , Elevation: 35 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-35	-37	-35	-24	-12	0	4	1	-8	-19	-28	-32	-19
Precip. mm	2	2	1	4	8	3	22	23	18	10	4	2	98

Nord, Greenland 81.5 degrees N , Elevation: 35 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-30	-30	-33	-23	-11	0	4	2	-8	-19	-24	-26	-16
Precip. mm	23	20	8	5	3	5	12	19	21	16	35	37	204

Polar Desert

Köppen Classification - **EF**.

Continental Arctic and Continental Polar air masses dominate.

Polar Desert climates are located in the high latitudes over continental areas, like Greenland and the Antarctica. This climate type covers a vast area of the Earth. For half of the year no solar radiation is received. During the summer months, available insolation is fairly high because of long days and a relatively transparent atmosphere. However, the ***albedo*** of snow-covered surfaces reflects up 90 percent of the ***insolation*** back to space. Average monthly temperatures are all generally below zero degrees Celsius. Winds are consistent and velocity is high enough to produce ***blizzard*** conditions most of the time.

Mirny, Antarctica 66.5 degrees S , Elevation: 30 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-2	-5	-10	-14	-16	-16	-17	-17	-17	-14	-7	-3	-12
Precip. mm	13	19	51	44	92	67	77	95	52	43	46	26	625

Plateau Station, Antarctica 79 degrees S , Elevation: 3625 m

{PRIVATE }	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.	Year
Temp. ° C	-34	-44	-57	-66	-66	-69	-68	-71	-65	-60	-44	-32	-56
Precip. mm	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

(w) Introduction to Applied Climatology

{PRIVATE}Urban Climatology

Urban and rural environments differ substantially in their micro-climate. These climatic differences are primarily caused by the alteration of the Earth's surface by human construction and the release of artificially created energy into the environment.

Energy Characteristics of Urban Areas

In a city, concrete, asphalt, and glass replace natural vegetation, and vertical surfaces of buildings are added to the normally flat natural rural landscape. Urban surfaces generally have a lower ***albedo***, greater heat ***conduction***, and more heat storage than the surfaces they replaced. The geometry of city buildings causes the absorption of a greater quantity of available incoming ***solar radiation*** and outgoing terrestrial ***infrared radiation***. Even in early morning and late afternoon the urban areas are intercepting and absorbing radiation on their vertical surfaces. In urban areas, large amounts of heat energy are added to the local energy balance through transportation, industrial activity, and the heating of buildings. In winter, the amount of heat generated from the burning of fossil fuels in New York City is 2.5 times greater than the heat absorbed from the sun. Finally, in rural areas, evaporation and transpiration from various natural surfaces act to cool the land surface and local atmosphere. In urban locations, drainage systems have been created to quickly remove surface water. Thus, little water is available for cooling.

Observed Climate of Cities

Urban areas tend to be warmer than the surrounding countryside. These differences in temperature are best observed at night under stable conditions when atmospheric mixing is at a minimum. Climatologists call this phenomenon the ***urban heat island***. The urban heat island is strongest at the city center where population densities are highest and industrial activity is at a maximum. The heat island effect has been described in many cities around the world, and temperature differences between city and country can be as high as 6 degrees Celsius.

Wind in urban areas is generally calmer than those in rural areas. This reduction in velocity is due the frictional effects of the city's vertical surfaces. However, some street and building configurations within a city can channel the wind and increase its velocity through a ***venturi*** effect. Certain parts of downtown Chicago and Winnipeg are noted for their unusually high wind speeds.

Climatologists have measured about up to 10 % more rainfall in urban areas. This increase may be due to the combined effect of particulate air pollution and increased convectional uplift. Air pollution may enhance rainfall by increasing the number of ***condensation nuclei*** through the atmospheric addition of smoke and dust particles. The additional generation of heat within the city increases the number of ***convection*** currents over that surface. Convection is required to initiate the development of ***thunderstorms***.

(x) Earth's Climatic History

Reconstructing Past Climates

A wide range of evidence exists to allow climatologists to reconstruct the Earth's past climate. This evidence can be grouped into three general categories.

The first category is **meteorological instrument records**. Common climatic elements measured by instruments include temperature, precipitation, wind speed, wind direction, and atmospheric pressure. However, many of these records are temporally quite short as many of the instruments used were only created and put into operation during the last few centuries or decades. Another problem with instrumental records is that large areas of the Earth are not monitored. Most of the instrumental records are for locations in populated areas of Europe and North America. Very few records exist for locations in less developed countries (LDCs), in areas with low human populations, and the Earth's oceans. Over the last half century many meteorological stations have been added in land areas previously not covered. Another important advancement in developing a global record of climate has been the recent use of remote satellites.

Written documentation and **descriptive accounts** of the weather make up the second general category of evidence for determining climate change. Weather phenomena commonly described in this type of data includes the prevailing character of the seasons of individual years, reports of floods, droughts, great frosts, periods of bitter cold, and heavy snowfalls. Large problems exist in the interpretation of this data because of its subjective nature.

Many types of **physical** and **biological data** can provide fossil evidence of the effects of fluctuations in the past weather of our planet. Scientists refer to this information as "***proxy data***" of past weather and climate. Examples of this type of data include tree ring width and density measurements, fossilized plant remains, insect and pollen frequencies in sediments, moraines and other glacial deposits, marine organism fossils, and the isotope ratios of various elements. Scientists using this type of data assume uniformity in the data record. Thus, the response measured from a physical or biological character existing today is equivalent to the response of the same character in the past. However, past responses of these characters may also be influenced by some other factor not accounted for. Some common examples of proxy data include:

Glacial Ice Deposits. Fluctuations in climate can be determined by the analysis of gas bubbles trapped in the ice which reflect the state of the atmosphere at the time they were deposited, the chemistry of the ice (concentration or ratio of major ions and isotopes of oxygen and hydrogen), and the physical properties of the ice.

Biological Marine Sediments. Climate change can be evaluated by the analysis of temporal changes in fossilized marine fauna and flora abundance, morpological changes in preserved organisms, coral deposits, and the oxygen isotopic concentration of marine organisms.

Inorganic Marine Sediments. This type of proxy data includes clay mineralogy, aeolian terrestial dust, and ice rafted debris.

Terrestrial Geomorphology and Geology Proxy Data. There are a number of different types of proxy data types in this group including glacial deposits, glacial erosional features, shoreline features, aeolian deposits, lake sediments, relict soil deposits, and speleothems (depositional features like stalactites and stalagmites).

Terrestial Biology Proxy Data. Variations in climate can be determined by the analysis of biological data like annual tree rings, fossilized pollen and other plant macrofossils, the abundance and distribution of insects and other organisms, and the biota in lake sediments.

Earth's Climatic History

Climatologists have used various techniques and evidence to reconstruct a history of the Earth's past climate. From this data they have found that during most of the Earth's history global temperatures were probably 8 to 15 degrees Celsius warmer than today. In the last billion years of climatic history, warmer conditions were broken by glacial periods starting at 925, 800, 680, 450, 330, and 2 million years before present.

The period from 2,000,000 - 14,000 B.P. (before present) is known as the ***Pleistocene*** or ***Ice Age***. During this period, large glacial ice sheets covered much of North America, Europe, and Asia for extended periods of time. The extent of the glacier ice during the Pleistocene was not static. The Pleistocene had periods when the glacier retreated (***interglacial***) because of warmer temperatures and advanced because of colder temperatures (***glacial***). During the coldest periods of the Ice Age, average global temperatures were probably 4 - 5 degrees Celsius colder than they are today.

The most recent glacial retreat is still going on. We call the temporal period of this retreat the ***Holocene epoch***. This warming of the Earth and subsequent glacial retreat began about 14,000 years ago (12,000 BC). The warming was shortly interrupted by a sudden cooling, known as the ***Younger-Dryas***, at about 10,000 - 8500 BC. Scientists speculate that this cooling may have been caused by the release of fresh water trapped behind ice on North America into the North Atlantic Ocean. The release altered vertical currents in the ocean which exchange heat energy with the atmosphere. The warming resumed by 8500 BC. By 5000 to 3000 BC average global temperatures reached their maximum level during the Holocene and were 1 to 2 degrees Celsius warmer than they are today. Climatologists call this period the ***Climatic Optimum***. During the climatic optimum many of the Earth's great ancient civilizations began and flourished. In Africa, the Nile River had three times its present volume, indicating a much larger tropical region.

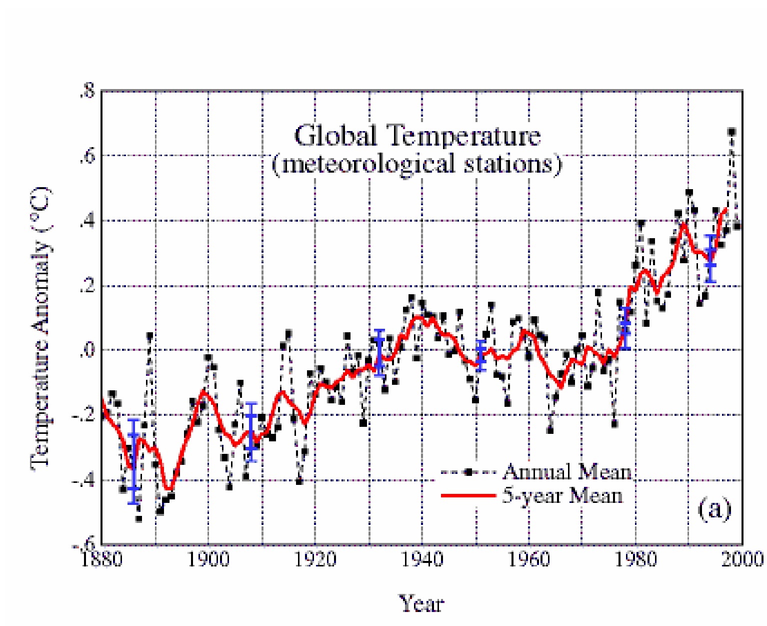
From 3000 to 2000 BC a cooling trend occurred. This cooling caused large drops in sea-level and the emergence of many islands (Bahamas) and coastal areas that are still above sea-level today. A short warming trend took place from 2000 to 1500 BC, followed once again by colder conditions. Colder temperatures from 1500 - 750 BC caused renewed ice growth in continental glaciers and alpine glaciers, and a sea-level drop of between 2 to 3 meters below present day levels.

The period from 750 BC - 800 AD saw warming up to 150 BC. Temperatures, however, did not get as warm as the Climatic Optimum. During the time of Roman Empire (150 BC - 300 AD) a cooling began that lasted until about 900 AD. At its height, the cooling caused the Nile River (829 AD) and the Black Sea (800-801 AD) to freeze.

The period 900 - 1200 AD has been called the ***Little Climatic Optimum***. It represents the warmest climate since the Climatic Optimum. During this period, the Vikings established settlements on Greenland and Iceland. The snow line in the Rocky Mountains was about 370 meters above current levels. A period of cool and more extreme weather followed the Little Climatic Optimum. A great drought in the American southwest occurred between 1276 and 1299. There are records of floods, great droughts and extreme seasonal climate fluctuations up to the 1400s.

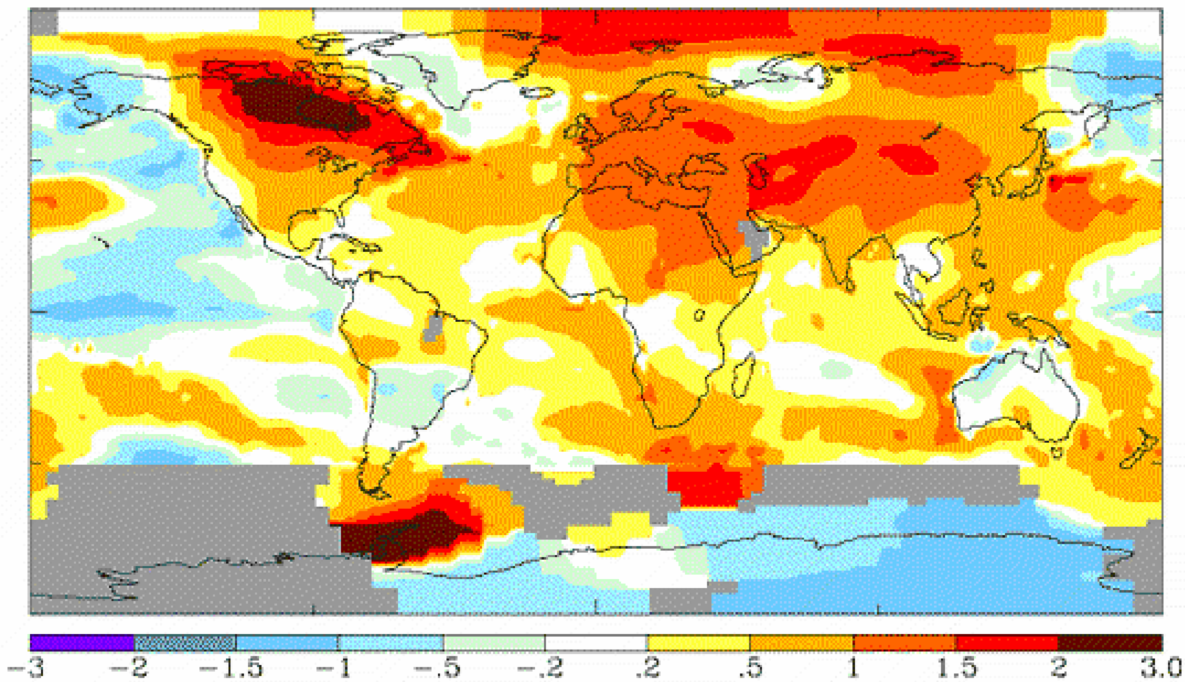
From 1550 to 1850 AD global temperatures were at their coldest since the beginning of the Holocene. Scientists call this period the ***Little Ice Age***. During the Little Ice Age, the average annual temperature of the Northern Hemisphere was about 1.0 degree Celsius lower than today. During the period 1580 to 1600, the western United States experienced one of its longest and most severe droughts in the last 500 years. Cold weather in Iceland from 1753 and 1759 caused 25 % of the population to die from crop failure and famine. Newspapers in New England were calling 1816 the year without a summer.

The period 1850 to present is one of general warming. **Figure 7x-1** describes the global temp erature trends from 1880 to 1999. This graph shows the yearly temperature anomalies that have occurred from an average global temperature calculated for the period. The graph indicates that the anomolies for the first 60 years of the record were consistently negative. However, beginning in 1935 positive anomolies became more common, and from 1980 to 1999 the anomolies were between 0.2 to 0.4 degrees Celsius higher that the average for the 119 year period of study.



{PRIVATE**}Figure 7x-1:** Near-global annual-mean temperature change for the period 1880 to 1999 (deviation from the normal period 1950-1980). (**Source:** NASA Goddard Institute for Space Studies - *Global Temperature Trends*).

In the 1930s and 1950s, the central United States experience two periods of extreme drought. The 1980s and 1990s had ten of the warmest years this century and possibly since the Little Climatic Optimum. Proxy and instrumental data indicate that 1998 was the warmest year globally in 1200 years of Earth history. In the following year, a La Nina developed and global temperatures dropped slightly. Nevertheless, the mean global temperatures recorded for this year was the sixth highest measurement since 1880. Many scientists believe the warmer temperatures of the 20th century are being caused by an enhancement of the Earth's ***greenhouse effect***.



{PRIVATE**}Figure 7x-2:** In 1999, most parts of the world were warmer than normal. The illustration above describes the annual temperature deviation (from the base period 1950-1980) in degrees Celsius for the Earth's surface. The illustration indicates that it was particularly warm across most of North America, northern Africa, and most of Eurasia. The tropical Pacific Ocean was cool due to a strong La Nina. (**Source:** NASA Goddard Institute for Space Studies - *Global Temperature Trends*).

(y) Causes of Climate Change

Figure 7y-1 illustrates the basic components that influence the state of the Earth's climatic system. Changes in the state of this system can occur externally (from extraterrestrial systems) or internally (from ocean, atmosphere and land systems) through any one of the described components. For example, an external change may involve a variation in the sun's output which would externally vary the amount of ***solar radiation*** received by the Earth's atmosphere and surface. Internal variations in the Earth's climatic system may be caused by changes in the concentrations of atmospheric gases, mountain building, volcanic activity, and changes in surface or atmospheric ***albedo***.

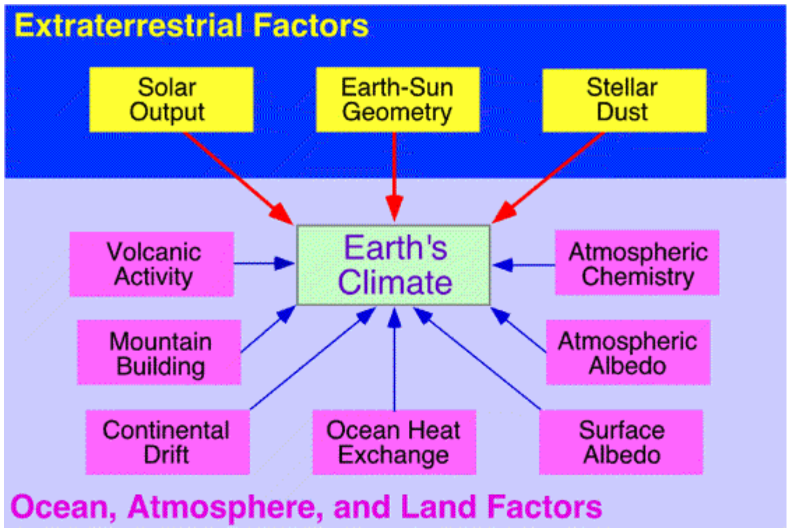


Figure 7y-1: Factors that influence the Earth's climate.

The work of climatologists has found evidence to suggest that only a limited number of factors are primarily responsible for most of the past episodes of climate change on the Earth. These factors include:

Variations in the Earth's orbital characteristics.

Atmospheric carbon dioxide variations.

Volcanic eruptions

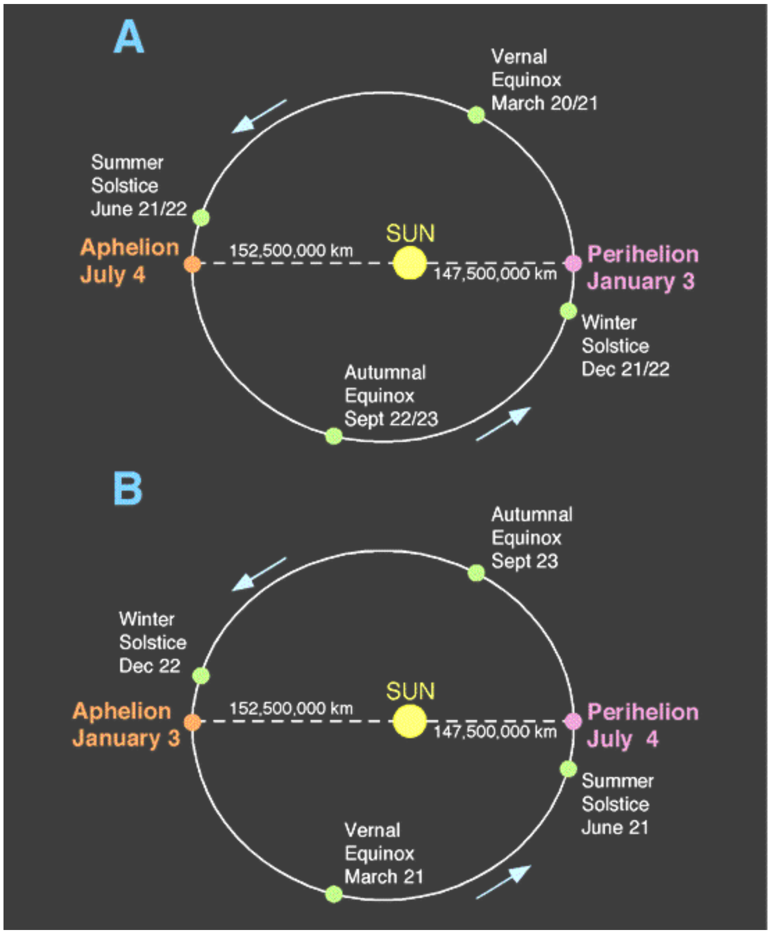
Variations in solar output.

Variations in the Earth's Orbital Characteristics

The ***Milankovitch theory*** suggests that normal cyclical variations in three of the Earth's orbital characteristics is probably responsible for some past climatic change. The basic idea behind this theory assumes that over time these three cyclic events vary the amount of ***solar radiation*** that is received on the Earth's surface.

The first cyclical variation, known as ***eccentricity***, controls the shape of the Earth's orbit around the sun. The orbit gradually changes from being elliptical to being nearly circular and then back to elliptical in a period of about 100,000 years. The greater the eccentricity of the orbit (i.e., the more elliptical it is), the greater the variation in solar energy received at the top of the atmosphere between the Earth's closest (***perihelion***) and farthest (***aphelion***) approach to the sun. Currently, the Earth is experiencing a period of low eccentricity. The difference in the Earth's distance from the sun between perihelion and aphelion (which is only about 3 %) is responsible for approximately a 7 % variation in the amount of solar energy received at the top of the atmosphere. When the difference in this distance is at its maximum (9 %), the difference in solar energy received is about 20 %.

The second cyclical variation results from the fact that, as the Earth rotates on its ***polar axis***, it wobbles like a spinning top changing the orbital timing of the ***equinoxes*** and ***solstices*** (see **Figure 7y-2** below). This effect is known as the ***precession of the equinox***. The **precession of the equinox** has a cycle of approximately 23,000 years. According to **illustration A**, the Earth is closer to the sun in January (***perihelion***) and farther away in July (***aphelion***) at the present time. Because of precession, the reverse will be true in 11,500 years and the Earth will then be closer to the sun in July (**illustration B**). This means, of course, that if everything else remains constant, 11,500 years from now seasonal variations in the Northern Hemisphere should be greater than at present (colder winters and warmer summers) because of the closer proximity of the Earth to the sun.



{PRIVATE}**Figure 7y-2:** Modification of the timing of aphelion and perihelion over time (**A** = today; **B** = 11,500 years into the future).

The third cyclical variation is related to the changes in the tilt (***obliquity***) of the Earth's axis of rotation over a 41,000 year period. During the 41,000 year cycle the tilt can deviate from approximately 22.5 to 24.5 degrees. At the present time, the tilt of the Earth's axis is 23.5 degrees. When the tilt is small there is less climatic variation between the summer and winter seasons in the middle and high latitudes. Winters tend to be milder and summers cooler. Warmer winters allow for more ***snow*** to fall in the high latitude regions. When the atmosphere is warmer it has a greater ability to hold water vapor and therefore more snow is produced at areas of ***frontal*** or ***orographic uplift***. Cooler summers cause snow and ice to accumulate on the Earth's surface because less of this frozen water is melted. Thus, the net effect of a smaller tilt would be more extensive formation of glaciers in the polar latitudes.

Periods of a larger tilt result in greater seasonal climatic variation in the middle and high latitudes. At these times, winters tend to be colder and summers warmer. Colder winters produce less snow because of lower atmospheric temperatures. As a result, less snow and ice accumulates on the ground surface. Moreover, the warmer summers produced by the larger tilt provide additional energy to melt and evaporate the snow that fell and accumulated during the winter months. In conclusion, glaciers in the polar regions should be generally receding, with other contributing factors constant, during this part of the **obliquity** cycle.

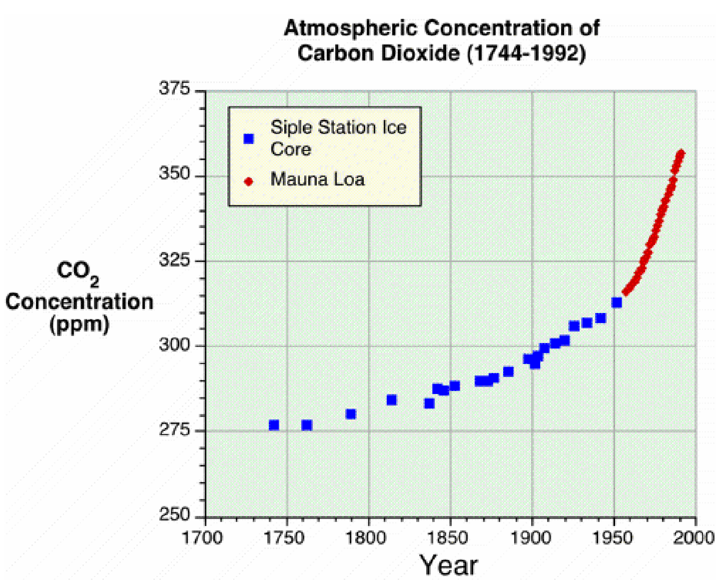
Computer models and historical evidence suggest that the Milankovitch cycles exert their greatest cooling and warming influence when the troughs and peaks of all three cycles coincide with each other.

Atmospheric Carbon Dioxide Variations

Studies of long term climate change have discovered a connection between the concentration of ***carbon dioxide*** in the atmosphere and mean global temperature. Carbon dioxide is one of the more important gases responsible for the ***greenhouse effect***. Certain atmospheric gases, like carbon dioxide, water vapor and ***methane***, are able to alter the energy balance of the Earth by being able to absorb ***longwave radiation*** emitted from the Earth's surface. The net result of this process and the re-emission of longwave back to the Earth's surface increases the quantity of heat energy in the Earth's climatic system. Without the greenhouse effect, the average global temperature of the Earth would be a cold -18 degrees Celsius rather than the present 15 degrees Celsius.

Researchers of the 1970s ***CLIMAP project*** found strong evidence in deep-ocean sediments of variations in the Earth's global temperature during the past several hundred thousand years of the Earth's history. Other subsequent studies have confirmed these findings and have discovered that these temperature variations were closely correlated to the concentration of carbon dioxide in the atmosphere and variations in solar radiation received by the planet as controlled by the Milankovitch cycles. Measurements indicated that atmospheric carbon dioxide levels were about 30 % lower during colder glacial periods. It was also theorized that the oceans were a major store of carbon dioxide and that they controlled the movement of this gas to and from the atmosphere. The amount of carbon dioxide that can be held in oceans is a function of temperature. Carbon dioxide is released from the oceans when global temperatures become warmer and diffuses into the ocean when temperatures are cooler. Initial changes in global temperature were triggered by changes in received solar radiation by the Earth through the Milankovitch cycles. The increase in carbon dioxide then amplified the global warming by enhancing the greenhouse effect.

Over the past three centuries, the concentration of carbon dioxide has been increasing in the Earth's atmosphere because of human influences (**Figure 7y-3**). Human activities like the burning of fossil fuels, conversion of natural prairie to farmland, and deforestation have caused the release of carbon dioxide into the atmosphere. From the early 1700s, carbon dioxide has increased from 280 parts per million to 360 parts per million in 1990. Many scientists believe that higher concentrations of carbon dioxide in the atmosphere will enhance the greenhouse effect making the planet warmer. Scientists believe we are already experiencing global warming due to an enhancement of the greenhouse effect. Most computer climate models suggest that the globe will warm up by 1.5 - 4.5 degrees Celsius if carbon dioxide reaches the predicted level of 600 parts per million by the year 2050.



{PRIVATE}Figure 7y-3: The following graph illustrates the rise in atmospheric carbon dioxide from 1744 to 1992. Note that the increase in carbon dioxide's concentration in the atmosphere has been **exponential** during the period examined. An extrapolation into the immediate future would suggest continued increases.

(**Source:** Neftel, A., H. Friedli, E. Moore, H. Lotscher, H. Oeschger, U. Siegenthaler, and B. Stauffer. 1994. Historical carbon dioxide record from the Siple Station ice core. pp. 11-14. In T.A. Boden, D.P. Kaiser, R.J. Sepanski, and F.W. Stoss (eds.) **Trends'93: A Compendium of Data on Global Change**. ORNL/CDIAC-65. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn. U.S.A. and Keeling, C.D., and T.P. Whorf. 1994. Atmospheric carbon dioxide records from sites in the SIO air sampling network. Pp. 16-26. In TA Boden, DP Kaiser, R.J. Sepanski, and F.W. Stoss (eds.) **Trends'93: A Compendium of Data on Global Change**. ORNL/CDIAC-65. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn. USA).

Volcanic Eruptions

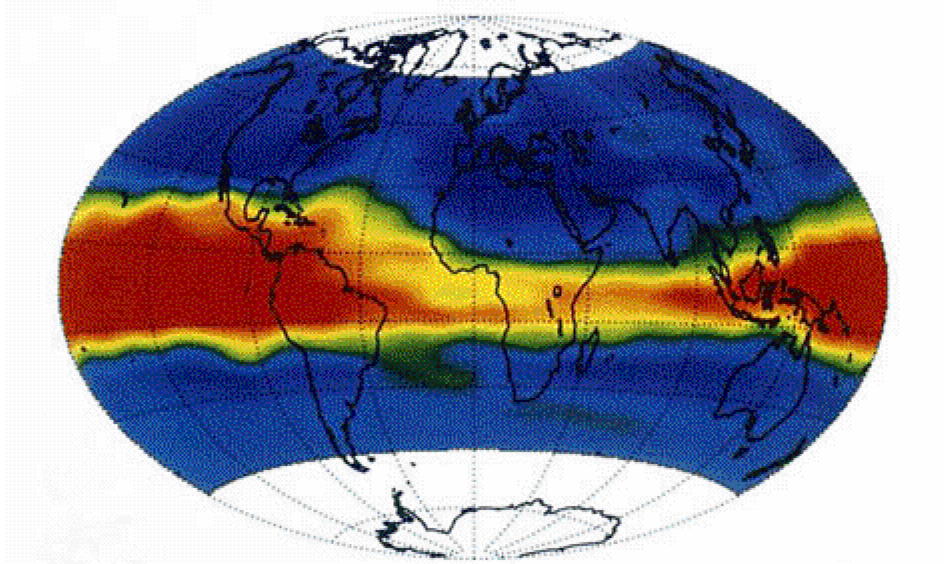
For many years, climatologists have noticed a connection between large explosive ***volcanic*** eruptions and short term climatic change (**Figure 7y-4**). For example, one of the coldest years in the last two centuries occurred the year following the Tambora volcanic eruption in 1815. Accounts of very cold weather were documented in the year following this eruption in a number of regions across the planet. Several other major volcanic events also show a pattern of cooler global temperatures lasting 1 to 3 years after their eruption.

{PRIVATE**}Figure 7y-4:** Explosive volcanic eruptions have been shown to have a short-term cooling effect on the atmosphere if they eject large quantities of sulfur dioxide into the stratosphere. This image shows the eruption of Mount St. Helens on May 18, 1980 which had a local effect on climate because of ash reducing the reception of solar radiation on the Earth's surface. Mount St. Helens had very minimal global effect on the climate because the eruption occurred at an oblique angle putting little sulfur dioxide into the stratosphere. (**Source:** U.S. Geological Survey, photograph by Austin Post).

At first, scientists thought that the dust emitted into the atmosphere from large volcanic eruptions was responsible for the cooling by partially blocking the transmission of ***solar radiation*** to the Earth's surface. However, measurements indicate that most of the dust thrown in the atmosphere returned to the Earth's surface within six months. Recent ***stratospheric*** data suggests that large explosive volcanic eruptions also eject large quantities of ***sulfur dioxide*** gas which remains in the atmosphere for as long as three years. Atmospheric chemists have determined that the ejected sulfur dioxide gas reacts with water vapor commonly found in the stratosphere to form a dense optically bright haze layer that reduces the atmospheric transmission of some of the sun's incoming radiation.

In the last century, two significant climate modifying eruptions have occurred. El Chichon in Mexico erupted in April of 1982, and Mount Pinatubo went off in the Philippines during June, 1991 (**Figure 7y-5**). Of these two volcanic events, Mount Pinatubo had a greater effect on the Earth's climate and ejected about 20 million tons of sulfur dioxide into the stratosphere (**Figure 7y-6**). Researchers believe that the Pinatubo eruption was primarily responsible for the 0.8 degree Celsius drop in global average air temperature in 1992. The global climatic effects of the eruption of Mount Pinatubo are believed to have peaked in late 1993. Satellite data confirmed the connection between the Mount Pinatubo eruption and the global temperature decrease in 1992 and 1993. The satellite data indicated that the sulfur dioxide plume from the eruption caused a several percent increase in the amount of sunlight reflected by the Earth's atmosphere back to space causing the surface of the planet to cool.

{PRIVATE**}Figure 7y-5:** Ash column generated by the eruption of Mount Pinatubo on June 12, 1991. The strongest eruption of Mount Pinatubo occurred three days later on June 15, 1991. (**Source:** US Geological Survey).	



{PRIVATE**}Figure 7y-6:** The following satellite image shows the distribution of Mount Pinatubo's sulfur dioxide and dust aerosol plume (red and yellow areas) between June 14 and July 26, 1991. Approximately 45 days after the eruption, the aerosol plume completely circled the Earth around the equator forming a band 20 to 50 degrees of latitude wide. Areas outside this band were clear of volcanic aerosols. Within a year, the sulfur dioxide continued to migrate towards the North and South Pole until it covered the entire Earth because of the dominant poleward flow of stratospheric winds (stratospheric winds circulate from the equator to the polar vortices at the North and South Poles). These observed patterns of aerosol movement suggest that tropical explosive volcanic eruptions probably have the greatest effect on the Earth's climate. Diffusion of aerosols by stratospheric winds from a tropical source results in the greatest latitudinal coverage of the sulfur dioxide across both the Northern and Southern Hemispheres. (**Source:** SAGE II Satellite Project - NASA).

Variations in Solar Output

Until recently, many scientists thought that the sun's output of radiation only varied by a fraction of a percent over many years. However, measurements made by satellites equipped with ***radiometers*** in the 1980s and 1990s suggested that the ***sun***'s energy output may be more variable than was once thought (**Figure 7y-7**). Measurements made during the early 1980s showed a decrease of 0.1 percent in the total amount of solar energy reaching the Earth over just an 18 month time period. If this trend were to extend over several decades, it could influence global climate. Numerical climatic models predict that a change in solar output of only 1 percent per century would alter the Earth's average temperature by between 0.5 to 1.0 degrees Celsius.

{PRIVATE**}Figure 7y-7:** The sun as seen at sunset. The sun is essentially the only source of energy for running the Earth's climate. Thus any change in its output will result in changes in the reception of insolation and the generation of heat energy which drives the climate system.

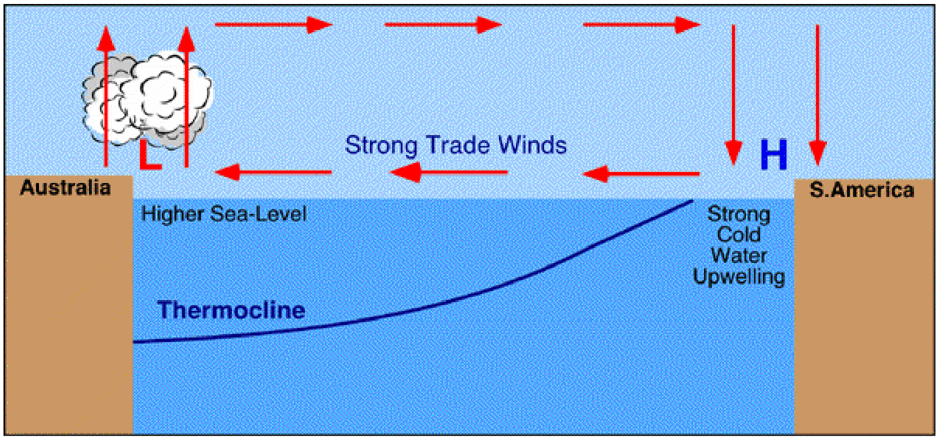
Scientists have long tried to also link ***sunspots*** to climatic change (also see the link ***www.sunspotcycle.com***). Sunspots are huge magnetic storms that are seen as dark (cooler) areas on the sun's surface. The number and size of sunspots show cyclical patterns, reaching a maximum about every 11, 90, and 180 years. The decrease in solar energy observed in the early 1980s correspond to a period of maximum sunspot activity based on the 11 year cycle. In addition, measurements made with a solar telescope from 1976 to 1980 showed that during this period, as the number and size of sunspots increased, the sun's surface cooled by about 6 degrees Celsius. Apparently, the sunspots prevented some of the sun's energy from leaving its surface. However, these findings tend to contradict observations made on longer times scales. Observations of the sun during the middle of the ***Little Ice Age*** (1650 to 1750) indicated that very little sunspot activity was occurring on the sun's surface. The **Little Ice Age** was a time of a much cooler global climate and some scientists correlate this occurrence with a reduction in solar activity over a period of 90 or 180 years. Measurements have shown that these 90 and 180 year cycles influence the amplitude of the 11 year sunspot cycle. It is hypothesized that during times of low amplitude, like the ***Maunder Minimum***, the sun's output of radiation is reduced. Observations by astronomers during this period (1645 to 1715) noticed very little sunspot activity occurring on the sun.

During periods of maximum sunspot activity, the sun's magnetic field is strong. When sunspot activity is low, the sun's magnetic field weakens. The magnetic field of the sun also reverses every 22 years, during a sunspot minimum. Some scientists believe that the periodic droughts on the Great Plains of the United States are in someway correlated with this 22 year cycle.

(z) El Nino, La Nina and the Southern Oscillation

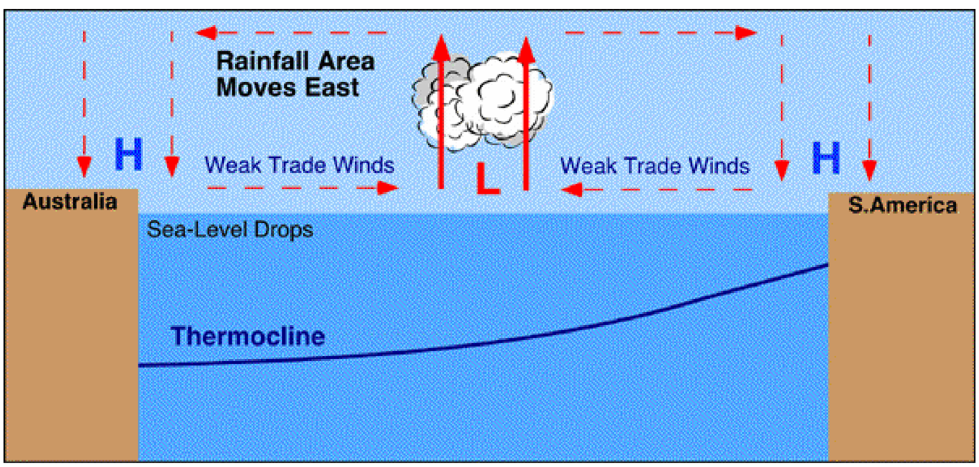
El Nino is the name given to the occasional development of warm ocean surface waters along the coast of Ecuador and Peru. When this warming occurs the usual upwelling of cold, nutrient rich deep ocean water is significantly reduced. El Nino normally occurs around Christmas and lasts usually for a few weeks to a few months. Sometimes an extremely warm event can develop that lasts for much longer time periods. In the 1990s, strong El Ninos developed in 1991 and lasted until 1995, and from fall 1997 to spring 1998.

The formation of an El Nino is linked with the cycling of a Pacific Ocean circulation pattern known as the ***southern oscillation***. In a normal year, a surface low pressure develops in the region of northern Australia and Indonesia and a high pressure system over the coast of Peru (see **Figure 7z-1** below). As a result, the ***trade winds*** over the Pacific Ocean move strongly from east to west. The easterly flow of the trade winds carries warm surface waters westward, bringing convective storms to Indonesia and coastal Australia. Along the coast of Peru, cold bottom water wells up to the surface to replace the warm water that is pulled to the west.

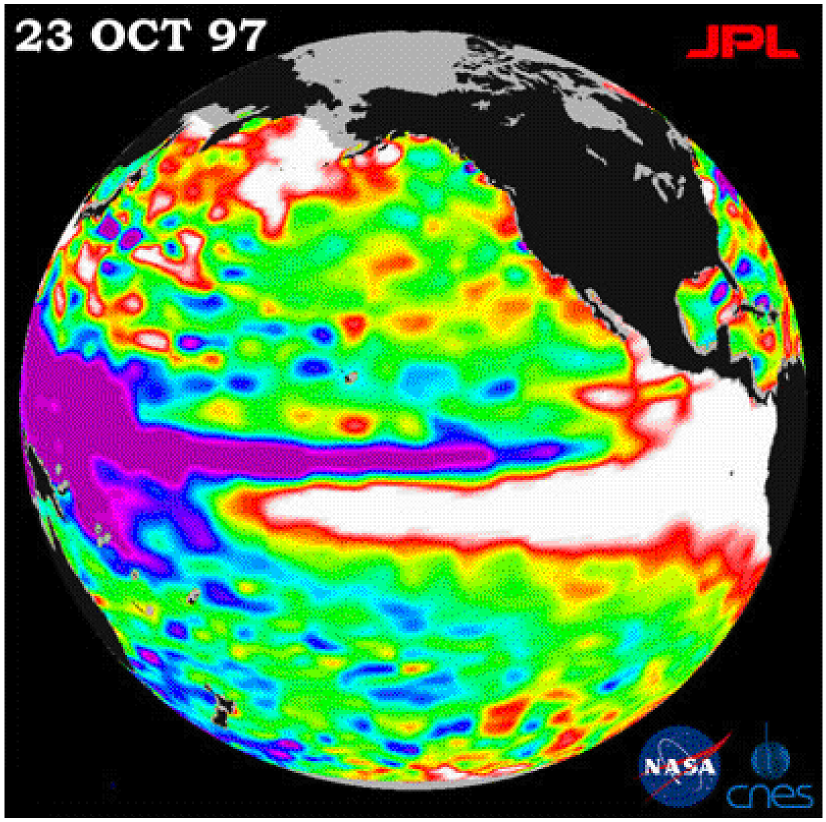


{PRIVATE**}Figure 7z-1 :** This cross-section of the Pacific ocean, along the equator, illustrates the pattern of atmospheric circulation typically found at the equatorial Pacific. Note the position of the ***thermocline***.

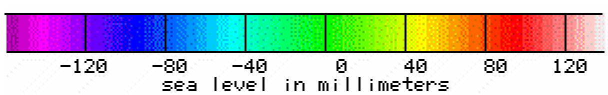
In an El Nino year, air pressure drops over large areas of the central Pacific and along the coast of South America (see **Figure 7z-2** below). The normal low pressure system is replaced by a weak high in the western Pacific (the ***southern oscillation***). This change in pressure pattern causes the trade winds to be reduced. This reduction allows the ***equatorial counter current*** (which flows west to east - see **ocean currents map** in **topic *8q***) to accumulate warm ocean water along the coastlines of Peru and Ecuador (**Figure 7z-3**). This accumulation of warm water causes the ***thermocline*** to drop in the eastern part of Pacific Ocean which cuts off the upwelling of cold deep ocean water along the coast of Peru. Climatically, the development of an El Nino brings drought to the western Pacific, rains to the equatorial coast of South America, and convective storms and hurricanes to the central Pacific.



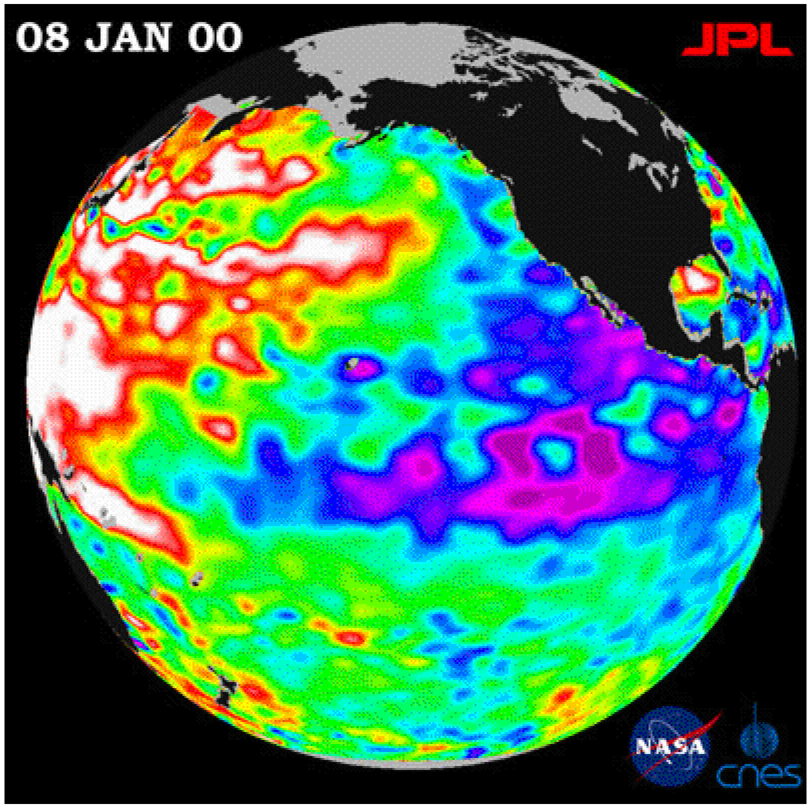
{PRIVATE**}Figure 7z-2 :** This cross-section of the Pacific ocean, along the equator, illustrates the pattern of atmospheric circulation that causes the formation of the El Nino. Note how position of the ***thermocline*** has changed from **Figure 7z-1**.



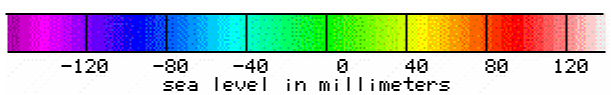
{PRIVATE**}Figure 7z-3:** NASA's TOPEX/Poseidon satellite is being used to monitor the presence of El Nino. Sensors on the satellite measure the height of the Pacific Ocean. The scale below describes the relationship between image color and the relative surface height of the ocean. In the image above, we can see the presence of a strong El Nino event in the eastern Pacific (October, 1997). The presence of the El Nino causes the height of the ocean along the equator to increase from the middle of the image to the coastline of Central and South America. (**Source:** NASA - *TOPEX/Poseidon*).



After an El Nino event weather conditions usually return back to normal. However, in some years the trade winds can become extremely strong and an abnormal accumulation of cold water can occur in the central and eastern Pacific (**Figure 7z-4**). This event is called a ***La Nina***. A strong La Nina occurred in 1988 and scientists believe that it may have been responsible for the summer drought over central North America. The most recent La Nina began developing in the middle of 1998 and have been persistent into the winter of 2000. During this period, the Atlantic ocean has seen very active ***hurricane*** seasons in 1998 and 1999. In 1998, ten ***tropical storm*** developed of which six become full-blown hurricanes. One of the hurricanes that developed, named Mitch, was the strongest October hurricane ever to develop in about 100 years of record keeping. Some of the other weather effects of La Nina include abnormally heavy ***monsoons*** in India and Southeast Asia, cool and wet winter weather in southeastern Africa, wet weather in eastern Australia, cold winter in western Canada and northwestern United States, winter drought in the southern United States, warm and wet weather in northeastern United States, and an extremely wet winter in southwestern Canada and northwestern United States.



{PRIVATE**}Figure 7z-4:** TOPEX/Poseidon satellite image of a moderate La Nina condition (January, 2000). The scale below describes the relationship between image color and the relative surface height of the ocean. The presence of the La Nina causes the height of the ocean either side of the equator to decrease from the middle of the image to the coastline of North, Central, and South America. (**Source:** NASA - *TOPEX/Poseidon*).



Prior to the 1980s and 1990s, strong El Nino events occurred on average every 10 to 20 years. In the early 1980s, the first of a series of strong events developed. The El Nino of 1982-83 brought extreme warming to the equatorial Pacific. Surface sea temperatures in some regions of the Pacific Ocean rose 6 degrees Celsius above normal. The warmer waters had a devastating effect on marine life existing off the coast of Peru and Ecuador. Fish catches off the coast of South America were 50 % lower than the previous year. The 1982-83 El Nino also had a pronounced influence on weather in the equatorial Pacific region and world wide. Severe droughts occurred in Australia, Indonesia, India and southern Africa. Dry conditions in Australia resulted in a 2 billion dollar loss in crops, and millions of sheep and cattle died from lack of water. Heavy rains were experienced in California, Ecuador, and the Gulf of Mexico.

Our understanding of the processes responsible for the development of El Nino is still incomplete. Scientists are able to predict the future development of an event by noting the occurrence of particular weather precursors. Researchers also now have a pretty complete understanding of the global weather effects caused by the formation of an El Nino (see **Figure 7z-5**).

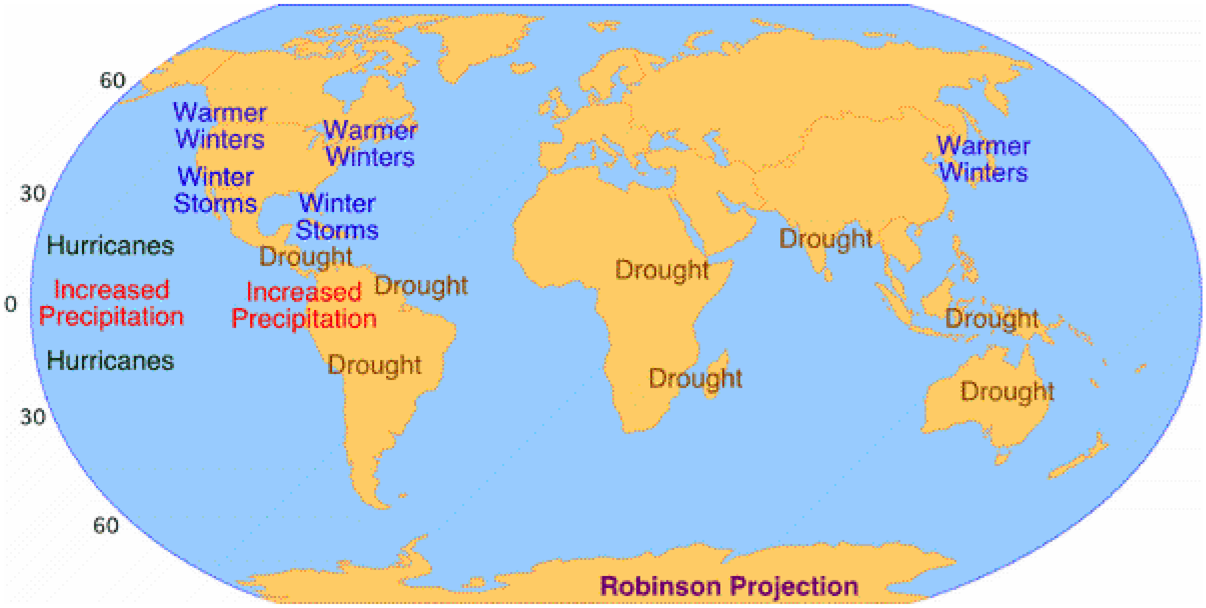


Figure 7z-5: Global climatological effects of the El Nino.

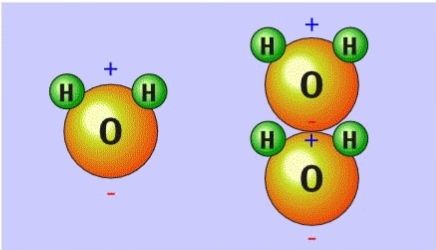
8) Introduction to Hydrology

(a) Physical Properties of Water

We live on a planet that is dominated by water. More than 70 % of the Earth's surface is covered with it. Scientists estimate that the ***hydrosphere*** contains about 1.36 billion cubic kilometers of this substance mostly in the form of a liquid that occupies topographic depressions on the Earth.

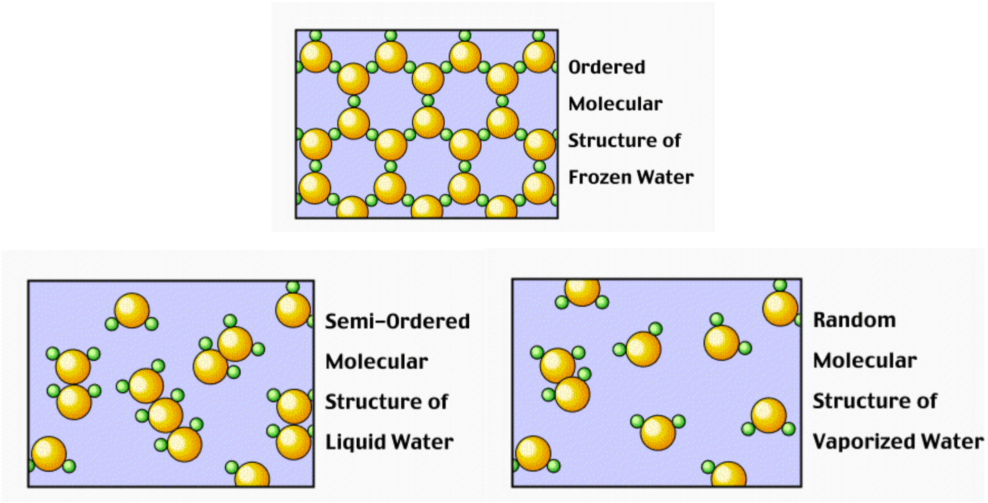
Water is also essential for life. Water is the major constituent of almost all life forms. Most animals and plants contain more than 60 % water by volume. Without water life would probably never have developed on our planet.

Water has a very simple atomic structure. This structure consists of two hydrogen atoms bonded to one oxygen atom (**Figure 8a-1**). The nature of the atomic structure of water causes its molecules to have unique electrochemical properties. The hydrogen side of the water molecule has a slight positive charge (see **Figure 8a-1**). On the other side of the molecule a negative charge exists. This molecular polarity causes water to be a powerful solvent and is responsible for its strong ***surface tension*** (for more information on these two properties see the discussion below).



{PRIVATE**}Figure 8a-1**: The atomic structure of a water molecule consists of two hydrogen (**H**) atoms joined to one oxygen (**O**) atom. The unique way in which the hydrogen atoms are attached to the oxygen atom causes one side of the molecule to have a negative charge and the area in the opposite direction to have a positive charge. The resulting polarity of charge causes molecules of water to be attracted to each other forming strong molecular bonds.

When water makes a physical ***phase change*** its molecules arrange themselves in distinctly different patterns (**Figure 8a-2**). The pattern taken by water when its is frozen causes its volume to expand and its density to decrease. Expansion of water at freezing allows ice to float on top of liquid water.



{PRIVATE}**Figure 8a-2**: The three diagrams above illustrate the distinct patterns of molecular arrangement in water when it changes its physical state from ice to water to gas. When water is frozen its molecules arrange themselves in a particular highly organized rigid geometric pattern that causes the mass of water to expand and to decrease in density. The diagram above shows a slice through a mass of ice that is one molecule wide. In the liquid phase, water molecules arrange themselves into small groups of joined particles. The fact that these arrangements are small allows liquid water to move and flow. Water in the form of a gas is highly charged with energy. This high energy state causes the molecules to be always moving reducing the likelihood of bonds between individual molecules from forming.

Water has several other unique physical properties. These properties are:

Water has a high ***specific heat***. Specific heat is the amount of energy required to change the temperature of a substance. Because water has a high specific heat, it can absorb large amounts of heat energy before it begins to get hot. It also means that water releases heat energy slowly when situations cause it to cool. Water's high specific heat allows for the moderation of the Earth's climate and helps organisms regulate their body temperature more effectively.

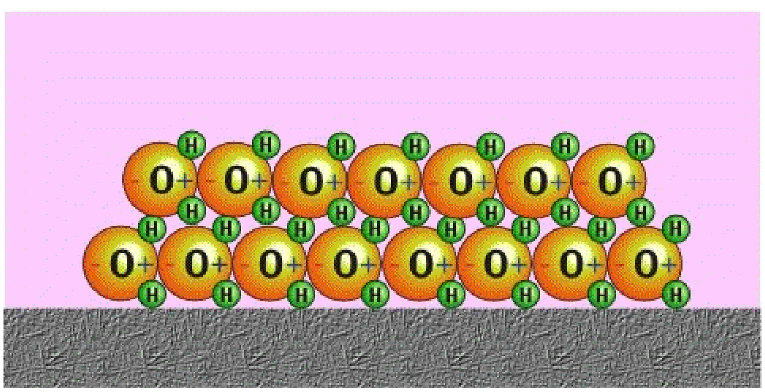
Water in a pure state has a neutral ***pH***. As a result, pure water is neither ***acidic*** nor ***basic***. Water changes its pH when substances are dissolved in it. Rain has a naturally acidic pH of about 5.6 because it contains natural derived carbon dioxide and sulfur dioxide.

Water ***conducts*** heat more easily than any liquid except mercury. This fact causes large bodies of liquid water like lakes and oceans to have essentially a uniform vertical temperature profile.

Water exists as a liquid over an important range of temperature from 0 - 100 degrees Celsius. This range allows water to remain as a liquid in most places on the Earth.

Liquid water is a universal **solvent**. It is able to dissolve a large number of different chemical compounds. This feature also enables water to carry solvent nutrients in ***runoff***, ***infiltration***, ***groundwater flow***, and living organisms.

Water has a high ***surface tension*** (**Figure 8a-3**). In other words, water is adhesive and elastic, and tends to aggregate in drops rather than spread out over a surface as a thin film. This phenomenon also causes water to stick to the sides of vertical structures despite gravity's downward pull. Water's high surface tension allows for the formation of water droplets and waves, allows plants to move water (and dissolved nutrients) from their roots to their leaves, and the movement of blood through tiny vessels in the bodies of some animals.



{PRIVATE}**Figure 8a-3**: The following illustration shows how the water molecules are attracted to each other to create high surface tension. This property can cause water to exist as an extensive thin film over solid surfaces. In the example above, the film is two layers of molecules thick.

Water is the only substance on Earth that exists in all three **physical states of matter**: **solid**, **liquid**, and **gas**. Incorporated in the changes of state are massive amounts of heat exchange. This feature plays an important role in the redistribution of heat energy in the Earth's atmosphere. In terms of heat being transferred into the atmosphere, approximately 3/4's of this process is accomplished by the evaporation and condensation of water.

The freezing of water causes it to expand. When water freezes it expands rapidly adding about 9 % by volume. Fresh water has a maximum density at around 4 degrees Celsius (see **Table 8a-1**). Water is the only substance on this planet where the maximum density of its mass does not occur when it becomes solidified.

Table 8a-1: Density of water at various temperatures.

{PRIVATE}**Temperature** (degrees Celsius)	**Density** (grams per cubic centimeter)
0 (solid)	0.9150
0 (liquid)	0.9999
4	1.0000
20	0.9982
40	0.9922
60	0.9832
80	0.9718
100 (gas)	0.0006

(b) The Hydrologic Cycle

The **hydrologic cycle** is a conceptual model that describes the storage and movement of water between the ***biosphere***, ***atmosphere***, ***lithosphere***, and the ***hydrosphere*** (see **Figure 8b-1**). Water on this planet can be stored in any one of the following reservoirs: ***atmosphere***, ***oceans***, ***lakes***, ***rivers***, ***soils***, ***glaciers***, ***snowfields***, and ***groundwater***.

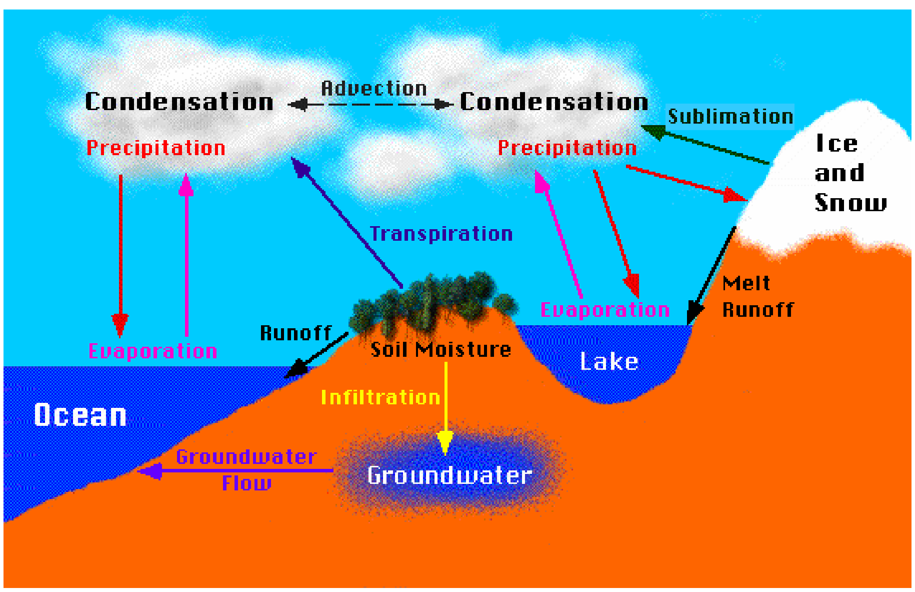


Figure 8b-1: Hydrologic Cycle.

Water moves from one reservoir to another by way of processes like ***evaporation***, ***condensation***, ***precipitation***, ***deposition***, ***runoff***, ***infiltration***, ***sublimation***, ***transpiration***, ***melting***, and ***groundwater flow***. The oceans supply most of the evaporated water found in the atmosphere. Of this evaporated water, only 91 % of it is returned to the ocean basins by way of precipitation. The remaining 9 % is transported to areas over landmasses where climatological factors induce the formation of precipitation. The resulting imbalance between rates of evaporation and precipitation over land and ocean is corrected by runoff and groundwater flow to the oceans.

The planetary water supply is dominated by the oceans (see **Table 8b-1**). Approximately 97 % of all the water on the Earth is in the oceans. The other 3 % is held as freshwater in glaciers and icecaps, groundwater, lakes, soil, the atmosphere, and within life.

Table 8b-1: Inventory of water at the Earth's surface.

{PRIVATE}Reservoir	Volume **(cubic km x 10,000,000)**	Percent of Total
Oceans	1370	97.25
Ice Caps and Glaciers	29	2.05
Groundwater	9.5	0.68
Lakes	0.125	0.01
Soil Moisture	0.065	0.005
Atmosphere	0.013	0.001
Streams and Rivers	0.0017	0.0001
Biosphere	0.0006	0.00004

Water is continually cycled between its various reservoirs. This cycling occurs through the processes of ***evaporation***, ***condensation***, ***precipitation***, ***deposition***, ***runoff***, ***infiltration***, ***sublimation***, ***transpiration***, ***melting***, and ***groundwater flow***. **Table 8b-2** describes the approximate residence times of water in the major reservoirs. On average water is renewed in rivers once every 16 days. Water in the atmosphere is completely replaced once every 8 days. Slower rates of replacement occur in large lakes, glaciers, ocean bodies and groundwater. Replacement in these reservoirs can take from hundreds to thousands of years. Some of these resources (especially groundwater) are being used by humans at rates that far exceed their renewal times. This type of resource use is making this type of water effectively ***nonrenewable.***

Table 8b-2: Approximate residence time of water found in various reserviors.

{PRIVATE}**Reservoir**	**Approximate Residence Time**
Glaciers	40 years
Seasonal Snow Cover	0.4 years
Soil Moisture	0.2 years
Groundwater: Shallow	200 years
Groundwater: Deep	10,000 years
Lakes	100 years
Rivers	0.04 years

(c) Atmospheric Humidity

Introduction

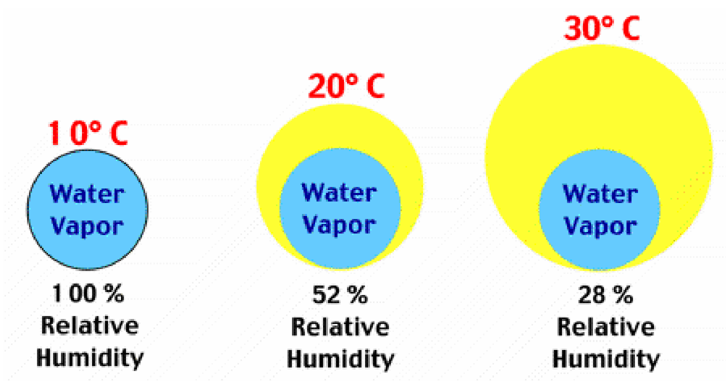
The term ***humidity*** is used in climatology and meteorology to describe the amount of water vapor held in the atmosphere. Scientists in these fields of study have developed a number of different measures of atmospheric humidity. We are primarily interested in two of these measures: ***saturation mixing ratio*** and ***relative humidity***.

There is a limit to how much water air can hold. However, this limit varies with temperature. Simply, warm air holds more water vapor than cold air. A more scientific description of this phenomenon is illustrated in the **Table 8c-1** with the ***saturation mixing ratio*** of water at various temperatures. This measure of humidity describes the maximum amount of water vapor air can hold at a given temperature in grams per kilogram of dry air. It is important to note that the relationship between temperature and water content in the air described in this table is not linear but exponential.

Table 8c-1: Saturation mixing ratio (at 1000 mb).

{PRIVATE}**Temperature Degrees Celsius**	**Vapor (g) per Kilogram of Dry Air**
50	88.12
40	49.81
30	27.69
20	14.85
10	7.76
0	3.84

The most commonly used measure of humidity is ***relative humidity***. Relative humidity can be simply defined as the amount of water in the air relative to the saturation amount the air can hold at a given temperature multiplied by 100. Air with a relative humidity of 50 % contains a half of the water vapor it could hold at a particular temperature. **Figure 8c-1** illustrates the concept of relative humidity.

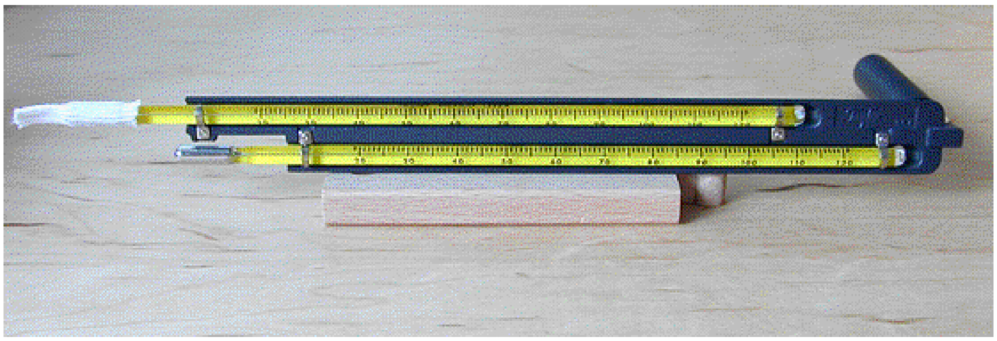


{PRIVATE}**Figure 8c-1:** The following illustration describes how relative humidity changes in a parcel of air with an increase in air temperature. At 10 degrees Celsius, a parcel of dry air weighing one kilogram can hold a maximum of 7.76 grams of water vapor (see **Table 8c-1**). In this state, the parcel of air would be at saturation and its ***relative humidity*** would be 100 %. Increasing the temperature of this parcel, without adding or removing any water, would increase its ability to hold water vapor. According to **Table 8c-1**, a 10 degree Celsius rise in temperature would increase the ***saturation mixing ratio*** of this parcel of air to 14.85 grams. Since no water has been added or removed, the actual amount of water in the parcel would remain 7.76 grams. This quantity is known as the **actual *mixing ratio***. Dividing the actual mixing ratio by the saturation mixing ratio and then multiplying this number by 100 determines the relative humidity of the parcel of air (7.76/14.85 x 100 = 52 %). At a temperature of 20 degrees Celsius, relative humidity would be 52 %. Raising the temperature of the parcel of air by another 10 degrees Celsius would again lower its relative humidity. In this state, the actual mixing ratio would still be 7.76 grams, while the saturation mixing ratio would increase to 27.69 grams. Relative humidity would drop to 28 % at a temperature of 30 degrees Celsius (7.76/27.69 x 100 = 28 %).

Measuring Humidity

Humidity can be measured using a variety of instruments. Relative humidity is often determined using a ***sling psychrometer*** or a ***hair hygrometer***. A ***sling psychrometer*** is a device that consists of two ***thermometers*** joined to a piece of plastic or metal (**Figure 8c-2**). One of the thermometers, called the ***wet-bulb thermometer***, has small cloth hood (wick) that is pulled over the reservoir bulb. The other thermometer has no hood and is called the ***dry-***

bulb thermometer. At one end of instrument is a rotating handle. To use the sling psychrometer, the wick is moistened with clean water and the device is twirled in the air using the handle. As the device is spun in the air, evaporation of the water from the wet-bulb thermometer occurs cooling it. The amount of evaporation and cooling taking place is controlled by the dryness of the air. If the air is saturated, the wet-bulb and dry-bulb thermometers would have the same temperature because no evaporation can occur. After a few minutes of twirling, the temperatures of the wet-bulb and dry-bulb thermometers are determined, a value called the ***wet-bulb depression*** is calculated (dry-bulb minus wet-bulb temperature), and a ***psychrometric table*** is used to find the corresponding relative humidity from the dry-bulb temperature and wet-bulb depression.



{PRIVATE**}Figure 8c-2:** Sling psychrometer. Note the wet-bulb thermometer is located on top.

Hair hygrometers work on the fact that hair changes its length when humidity varies. This device usually consists of a number of human or horse hairs connected to a mechanical lever system. When humidity increases the length of the hairs becomes longer. This change in length is then transmitted and magnified by the lever system into a measurement of relative humidity.

Humidity is also measured on a global scale using remotely placed satellites (**Figure 8c-3**). These satellites are able to detect the concentration of water in the ***troposphere*** at altitudes between 4 and 12 kilometers. Satellites that can measure water vapor have sensors that are sensitive to infrared radiation. Water vapor specifically absorbs and re-radiates radiation in this spectral band. Satellite water vapor imagery plays an important role in monitoring climate conditions (like the formation of ***thunderstorms***) and in the development of future weather forecasts.

Dew Point and Frost Point

Associated with relative humidity is ***dew point*** (if the dew point is below freezing, it is referred to as the ***frost point***). Dew point is the temperature at which water vapor saturates from an air mass into liquid or solid usually forming rain, snow, frost, or dew. Dew point normally occurs when a mass of air has a relative humidity of 100 %. This happens in the atmosphere as a result of cooling through a number of different processes.

(d) Condensation, Freezing, and Deposition

We have learned that water is available on the Earth in the following three forms: vapor; liquid; and solid. The process of water moving from one of these forms to another is called a **phase** or ***state change***. In the atmosphere, three processes act to create water droplets or ice crystals. These three processes are:

Condensation - water moving from a vapor to a liquid state.

Freezing - water moving from a liquid to a solid state.

Deposition - water moving from a vapor to a solid state.

For a phase change to occur heat energy must be added to or removed from water molecules. The formation of water droplets and ice crystals takes place when the water in the atmosphere is cooled. As air containing water vapor cools, the ***relative humidity*** of the air parcel increases until the dew or frost point is reached. At dew point (relative humidity = 100 %) water begins to condense into droplets. If 100 % relative humidity is reached below 0 degrees Celsius deposition occurs and ice crystals form.

Formation of water droplets and ice crystals also requires a surface for condensation, freezing, or deposition. In the atmosphere, these surfaces are microscopic particles of dust, smoke, and salt commonly called ***condensation nuclei***. ***Deposition nuclei***, six sided particles, are needed for the formation of ice crystals.

If there is a deficiency of nuclei, ***super-saturation*** can result and condensation, freezing, or deposition can only occur with a relative humidity that is greater than 100 %

(e) Cloud Formation Processes

Condensation or ***deposition*** of water above the Earth's surface creates clouds. In general, clouds develop in any air mass that becomes ***saturated*** (***relative humidity*** becomes 100 %). Saturation can occur by way of atmospheric mechanisms that cause the temperature of an air mass to be cooled to its ***dew point*** or ***frost point***. The following mechanisms or processes can achieve this outcome causing clouds to develop:

(1). ***Orographic uplift*** occurs when air is forced to rise because of the physical presence of elevated land. As the parcel rises it cools as a result of adiabatic expansion at a rate of approximately 10 degrees Celsius per 1000 meters until ***saturation***. The development of clouds and resulting heavy quantities of precipitation along the west coast of Canada are mainly due to this process.

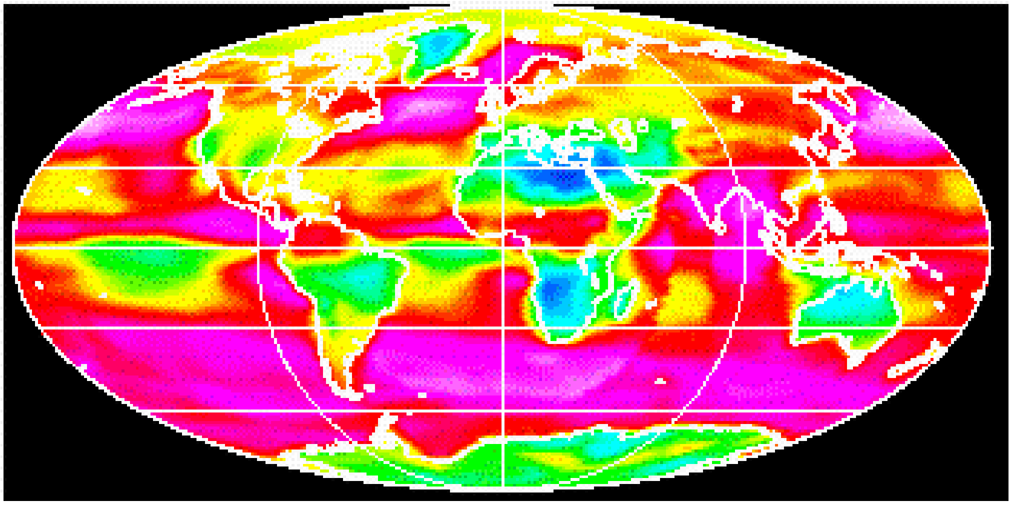
(2). ***Convectional lifting*** is associated with surface heating of the air at the ground surface. If enough heating occurs, the mass of air becomes warmer and lighter than the air in the surrounding environment, and just like a hot air balloon it begins to rise, expand, and cool. When sufficient cooling has taken place ***saturation*** occurs forming clouds. This process is active in the interior of continents and near the equator forming ***cumulus clouds*** and or cumulonimbus clouds (***thunderstorms***). The ***rain*** that is associated with the development of ***thunderstorm*** clouds is delivered in large amounts over short periods of time in extremely localized areas.

(3). ***Convergence*** or *..\physgeoglos/c.html - anchor432530*___frontal lifting___ takes place when two ***masses of air*** come together. In most cases, the two air masses have different temperature and moisture characteristics. Oneof the air masses is usually warm and moist, while the other is cold and dry. The leading edge of the latter air mass acts as an inclined wall or front causing the moist warm air to be lifted. Of course the lifting causes the warm moist air mass to cool due to expansion resulting in ***saturation***. This cloud formation mechanism is common at the mid-latitudes where ***cyclones*** form along the ***polar front*** and near the equator where the ***trade winds*** meet at the ***intertropical convergence zone***.

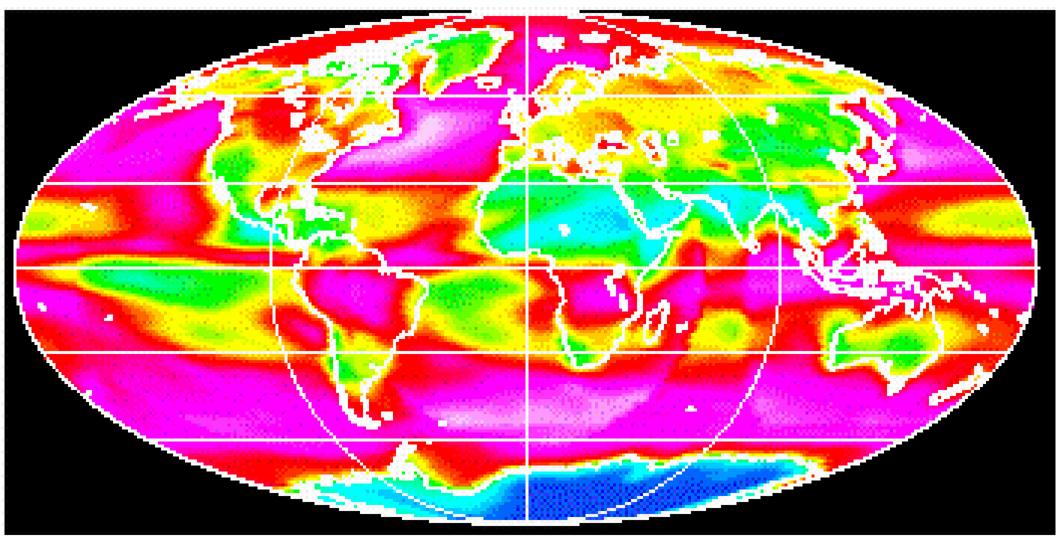
(4). **Radiative cooling** occurs when the sun is no longer supplying the ground and overlying air with energy derived from solar insolation (e.g., night). Instead, the surface of the Earth now begins to lose energy in the form of longwave radiation which causes the ground and air above it to cool. The clouds that result from this type of cooling take the form of surface ***fog***.

Of course these causes of cloud development do not always act in a singular fashion. It is possible to get combinations of all four types, such as when convection and orographic uplift cause summer afternoon cloud development and showers in the mountains.

The following two images (**Figures 8e-1** and **8e-2**) describe percent global cloud coverage averaged for the months of July and January using 8 years of data. From these images one can see the effects of ***convergence***, ***frontal lifting***, and ***orographic uplift*** in enhancing cloud cover over select areas of the Earth.



{PRIVATE**}Figure 8e-1:** Percent cloud cover: July 1983-1990. Highest levels of cloud cover occur over the mid-latitude cyclone storm tracks of both hemispheres, Intertropical Convergence Zone over land surfaces, and the Indian Monsoon region (orographic lifting). Lowest values occur over the subtropical deserts, the subsidence regions of the subtropical oceans, and the polar regions. Color range: blue - red - white, Values: 0 - 100%. Global mean = 59%, Minimum = 1%, Maximum = 95%. (**Source:** NASA *Surface Radiation Budget Project*).



{PRIVATE**}Figure 8e-2:** Percent cloud cover: January 1984-1991. Highest levels of cloud cover occur over the mid-latitude cyclone storm tracks of both hemispheres and the Intertropical Convergence Zone over land surfaces. Lowest values occur over the subtropical deserts, the subsidence regions of the subtropical oceans, and over the South Pole. Color range: blue - red - white, Values: 0 - 100%. Global mean = 59%, Minimum = 1%, Maximum = 96%. (**Source:** NASA *Surface Radiation Budget Project*).

(f) Precipitation and Fog

Precipitation

We can define ***precipitation*** as any aqueous deposit, in liquid or solid form, that develops in a ***saturated*** atmosphere (***relative humidity*** equals 100 %) and falls to the ground generally from douds. Most clouds, however, do not produce precipitation. In many clouds, water droplets and ice crystals are too small to overcome natural **updrafts** found in the atmosphere. As a result, the tiny water droplets and ice crystals remain suspended in the atmosphere as clouds.

Water droplets and ice crystals can only fall to the Earth's surface if they grow to a size that can overcome **updrafts**. Conditions for growth can develop in clouds via two different processes.

In clouds with temperatures above freezing, turbulent atmospheric mixing can cause droplets to grow through **collision** and **coalescence**. One initial condition, however, must be met for this process to begin: droplet size in the cloud must be variable. This initial condition allows larger and heavier droplets to collide and coalesce with lighter smaller droplets during **downdraft** periods. If enough atmospheric mixing occurs the larger droplets can expand by up to 250 times and can become heavy enough to fall to the Earth's surface.

The other mechanism of precipitation development involves clouds whose temperature is below freezing. In these clouds, large ice crystals grow due to the differences in ***vapor pressure*** between ice crystals and ***supercooled*** water droplets. Vapor pressure differences between ice and supercooled water causes a net migration of water vapor from water droplets to ice crystals. The ice crystal then absorbs the water vapor, depositing it on their surface. At the same time, the loss of vapor from the water droplets causes them to shrink in size. A necessary initial requirement for this process is the presence of both ***condensation nuclei*** and ***deposition nuclei***. While deposition nuclei form ice crystals at temperatures just below zero degrees Celsius, condensation nuclei can remain liquid (***supercooled***) to temperatures as low as - 40 degrees Celsius depending on size. Because of this phenomenon, cold clouds can contain both ice crystals and supercooled water droplets. The relative proportion of these two types of particles determines whether snow crystals grow to a size to overcome atmospheric updrafts.

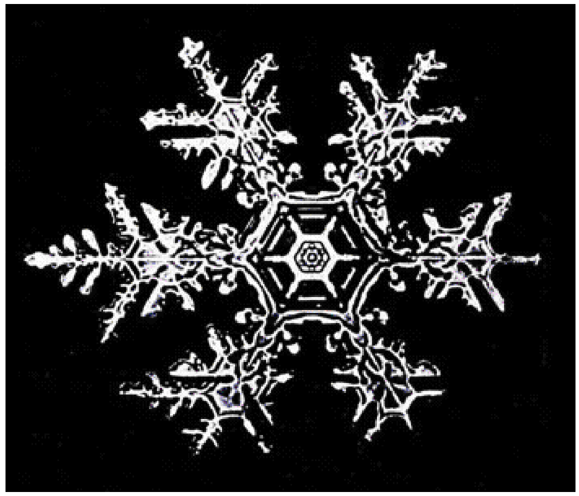
The following list describes the various types of precipitation that can form in the atmosphere:

Rain is any liquid deposit that falls from the atmosphere to the surface and has a diameter greater than 0.5 millimeters.

Freezing rain occurs when water droplets hit a cold surface and freeze. For this to occur a surface ***temperature inversion*** is required. In such an inversion, the surface must have a temperature below freezing, while the temperature of the atmosphere where the precipitation forms is above freezing. Surface inversions may develop from a variety of causes, but typically, they occur near the leading edge of cold air from the North as it pushes southward in the Northern Hemisphere (in the Southern Hemisphere surface inversions occur near the leading edge of cold air from the South as it pushes northward).

Ice pellets or ***sleet*** are transparent or translucent bits of frozen water with a diameter less than 5 millimeters. To form, these pellets require an environment where raindrops develop in an atmosphere where the temperature is above freezing and then fall into a lower layer of air with temperatures below freezing. In the lower layer of cold air the raindrops freeze into small ice pellets. Normally, to provide temperatures above freezing in the atmosphere above a colder atmospheric layer an upper air ***temperature inversion*** is required.

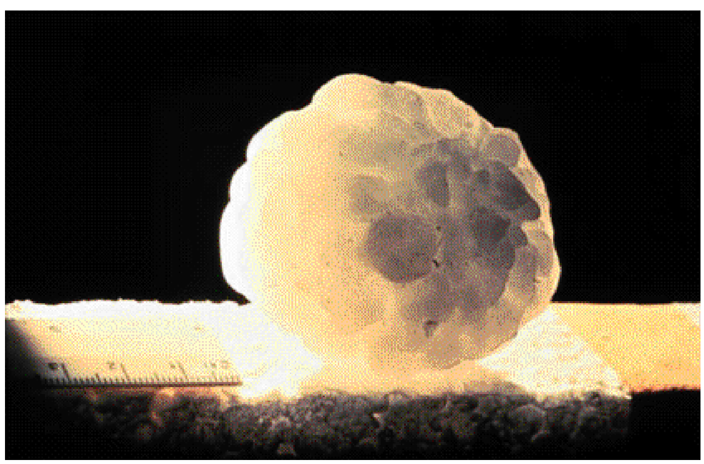
Snow can develop only from water vapor that deposits directly as a solid on a ***deposition nuclei*** at temperatures below freezing by passing the liquid state. A snowflake forms first as a very tiny crystal developing on a six-sided hexagonal deposition nuclei (**Figure 8f-1**). The ice crystal then grows fastest at the six points as these areas are more directly exposed to the atmosphere's water vapor. Snow is most common in winter just North of the center of ***mid-latitude cyclones*** in the Northern Hemisphere (south of a mid-latitude in the Southern Hemisphere). As the warm moist air travels around the center of lowest pressure, it overrides colder air located North of the low and is cooled to its saturation temperature, producing rainfall and snow. Snow generally occurs with East winds, since the winds at locations north of a mid-latitude cyclone are from the East.



{PRIVATE}**Figure 8f-1:** Close-up photograph of a snowflake. (**Source**: *NOAA Photo Collection Website*)**.**

Snow pellets are white, spherical grains of ice 2 to 5 millimeters in diameter. They can be distinguished from packed snowflakes since snow pellets are firm enough to bounce when they hit the ground. Snow pellets develop as ***supercooled*** droplets freeze on ice crystals. They may fall for a brief period as the precipitation changes from ***ice pellets*** to ***snow***.

Hail is a destructive form of precipitation that is 5 to 190 millimeters in diameter (**Figure 8f-2**). The large downdrafts in mature ***thunderstorm*** clouds provide the mechanism for hail formation. Hailstones normally have concentric shells of ice alternating between those with a milky appearance and those that are clear. The milky white shells, containing bubbles and partially melted snowflakes, correspond to a period of rapid freezing, while the clear shells develop as the liquid water freezes much more slowly.



{PRIVATE}**Figure 8f-2:** Hailstone measuring 21 centimeters in diameter. (**Source**: *NOAA Photo Collection Website*)**.**

Fog

If the air near the ground is cooled sufficiently, it becomes saturated and ***fog*** can develop. By definition, fog exists if the atmospheric visibility near the Earth's surface is reduced to 1 kilometer or less. Fogs composed primarily of water droplets can be classified according to the process that causes the air to cool and ***saturate***:

Radiation fog or **ground fog**, is generated as the Earth's surface cools by radiation loss at night. This type of fog is normally quite shallow.

Upslope fog is generated by airflow over topographic barriers. As the air is forced to rise upward where the atmospheric pressure is less, it is cooled by expansion and produces fog on the windward slopes of hills or mountains.

Advection fog may be generated by winds that contrast in temperature with the Earth's surface. Warm air advection can produce fog through contact cooling with a cold surface.

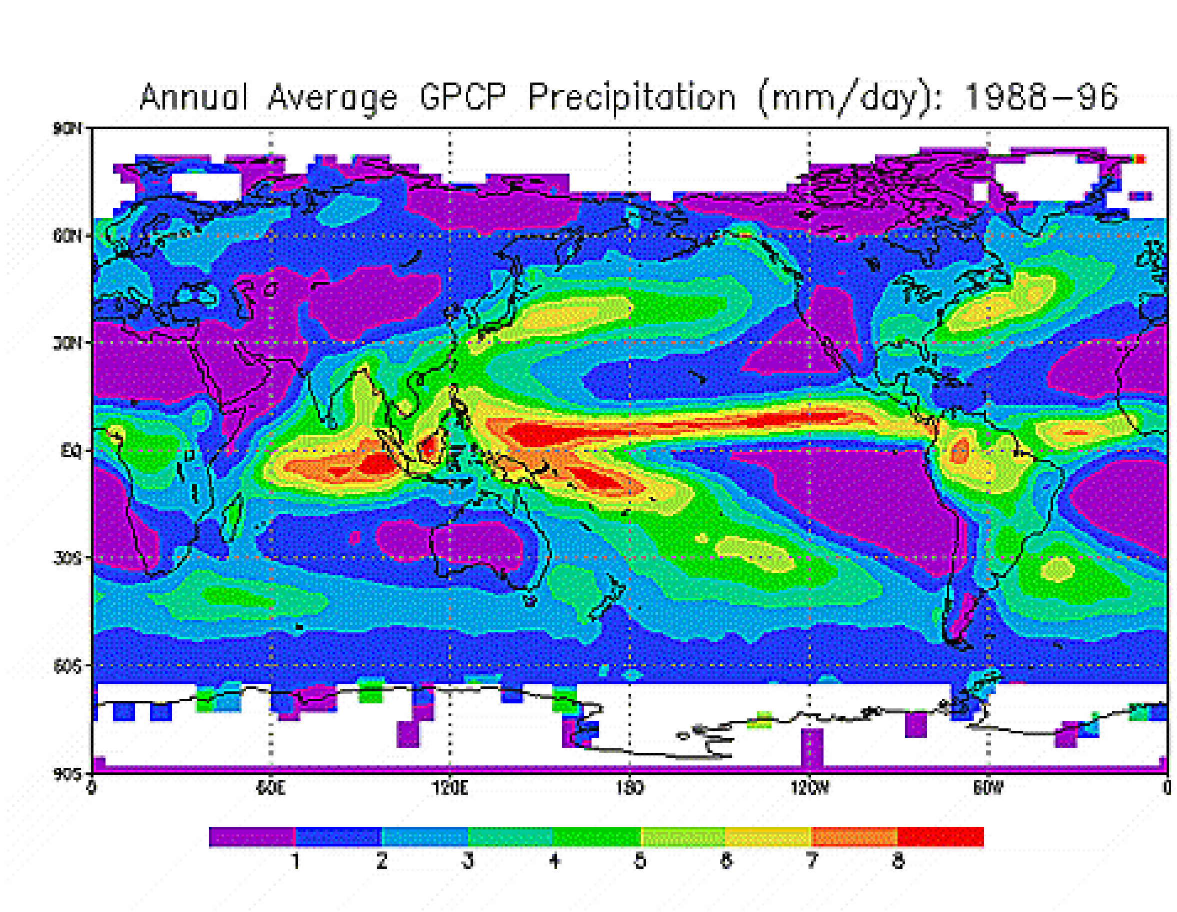
Evaporation fog occurs when you get cold air advection over warm water or warm, moist land surfaces resulting in fog formation as water evaporates into the cold air (**Figure 8f-3**). This type of fog is sometimes called **steam fog** or **sea smoke**.

{PRIVATE**}Figure 8f-3:** Evaportation fog forming over a lake as cool air flows over the warm water. (**Source**: *NOAA Photo Collection Website*)**.**

Frontal fog is produced as weather fronts, especially ***warm fronts***, pass through an area. Precipitation falling into the colder air ahead of the warm front may evaporate enough water to cause the formation of small droplets as fog near the ground.

(g) Global Distribution of Precipitation

The **Global Precipitation Climatology Project** (**GPCP**) was established by the **World Climate Research Program** (**WCRP**) in 1986 with the goal of providing monthly mean precipitation data on a 2.5 x 2.5 degree latitude-longitude grid for the period 1986-2000. The **GPCP** will accomplish this by combining infrared and microwave satellite estimates of precipitation with rain gauge data from more than 30,000 stations. Infrared precipitation measurements are obtained from **GOES** (United States), **GMS** (Japan) and **Meteosat** (European Community) geostationary satellites and **National Oceanic and Atmospheric Administration** (**NOAA**) operational polar orbiting satellites. Microwave estimates are obtained from the U.S. **Defense Meteorological Satellite Program** (**DMSP**) satellites using the **Special Sensor Microwave Imager** (**SSM/I**). Together these data sets will be used to validate general circulation and climate models, study the global hydrological cycle and diagnose the variability of the global climate system. **Figure 8g-1** describes mean annual global precipitation over an eight year period measured in millimeters per day.



{PRIVATE**}Figure 8g-1:** Mean annual global precipitation 1988-1996. (**Source:** NOAA *Global Precipitation Climatology Project*).

The average annual precipitation of the world is estimated to be 1050 millimeters per year or 2.9 millimeters per day. However, **Figure 8g-1** indicates that actual values vary from a minimum of 0 millimeters per day or to a maximum of 10 millimeters per day depending on location. The reasons for these patterns are as follows:

The deserts in the subtropical regions occur because these areas do not contain any mechanism for lifting air masses. In fact, these areas are dominated by subsiding air that results from global circulation patterns.

Continental areas tend to be dry because of their distance from moisture sources.

Polar areas are dry because cold air cannot hold as much moisture as warm air.

Areas near the equator achieve high rainfall amounts because constant solar heating encourages convection, and global circulation patterns cause northern and southern air masses to converge here causing ***frontal lifting***.

Mid-latitudes experience ***cyclonic*** activity and frontal lifting when polar and subtropical air masses meet at the polar front. Further, the air masses in this region generally move from West to East, causing levels of precipitation to decrease East of source regions.

Mountain ranges near water sources can receive high rainfalls because of ***orographic uplift***, if and only if the prevailing winds are in their favor. This can also result in a sharp reduction in rainfall in regions adjacent or on the ***leeward*** slopes of these areas. This phenomenon is commonly know as the ***rainshadow effect***.

Table 8g-1 describes some of the precipitation extremes recorded around the world.

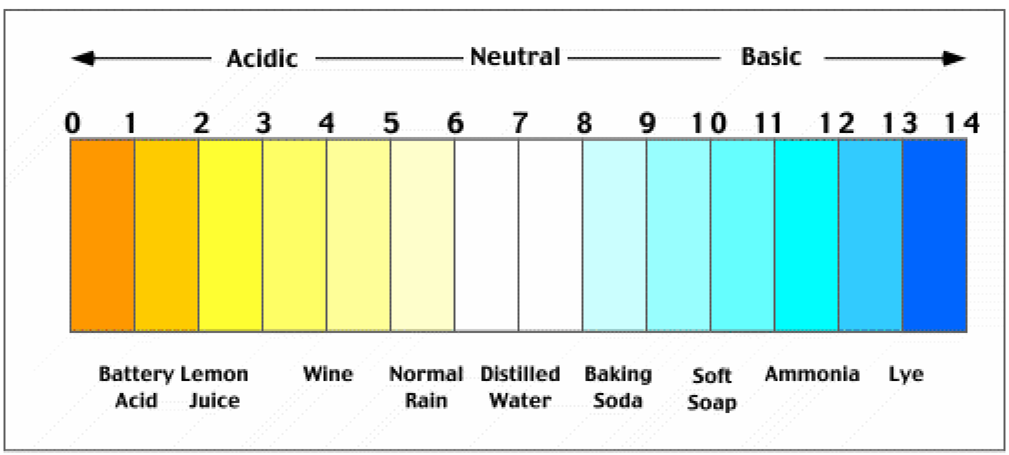
Table 8g-1: Precipitation extreme weather records.

{PRIVATE}**Record**	**Location**	**Amount (mm)**	**Date**
1-year Rainfall	Cherrapundi, India	26,470	1861
1-month Rainfall	Cherrapundi, India	9300	1861 (July)
Average Annual Rainfall	Mt. Waialeale, Hawaii, USA	11,680	
24 hr. Rainfall	Belouve, La Reunion Island	1350	Feb 28, 1964
Lowest Annual Average Rainfall	Arica, Chile	0.8	
Greatest 1 Month Snowfall	Tamarack, California, USA	9910	1911 (Jan)
Greatest Snowfall Single Storm	Mt. Shasta, California, USA	4800	Feb 13-19, 1959

(h) Acid Precipitation

Introduction

Acidic pollutants can be deposited from the atmosphere to the Earth's surface in wet and dry forms. The common term to describe this process is ***acid deposition***. The term ***acid precipitation*** is used to specifically describe wet forms of acid pollution that can be found in rain, sleet, snow, fog, and cloud vapor. An acid can be defined as any substance that when dissolved in water dissociates to yield corrosive hydrogen ions. The acidity of substances dissolved in water is commonly measured in terms of ***pH*** (defined as the negative logarithm of the concentration of hydrogen ions). According to this measurement scale solutions with pHs less than 7 are described as being ***acidic***, while a pH greater than 7.0 is considered ***alkaline*** (**Figure 8h-1**). Precipitation normally has a pH between 5.0 to 5.6 because of natural atmospheric reactions involving carbon dioxide. For comparison, distilled water, pure of any other stubstances, would have a pH of 7.0. Precipitation is considered to be acidic when its pH falls below 5.6 (which is 25 times more acidic than pure distilled water). Some sites in eastern North America have precipitation events with pHs as low as 2.3 or about 1000 times more acidic than natural.



{PRIVATE}**Figure 8h-1:** The pH scale. A value of 7.0 is considered neutral. Values higher than 7.0 are increasingly alkaline or basic. Values lower than 7.0 are increasingly acidic. The illustration above also describes the pH of some common substances.

Acid deposition is not a recent phenomenon. In the 17th century, scientists noted the ill effects that industry and acidic pollution was having on vegetation and people. However, the term ***acid rain*** was first used two centuries later when Angus Smith published a book called '**Acid Rain**' in 1872. In the 1960s, the problems associated with acid deposition became an international problem when fishermen noticed declines in fish numbers and diversity in many lakes throughout North America and Europe.

Acid Deposition Formation

Acid deposition can form as a result of two processes. In some cases, hydrochloric acid can be expelled directly into the atmosphere. More commonly it is due to ***secondary pollutants*** that form from the oxidation of ***nitrogen oxides*** (NO_x) or ***sulfur dioxide*** (SO_2) gases that are released into the atmosphere (see **Figure 8h-2**). Reactions at the Earth's surface or within the atmosphere can convert these pollutants into ***nitric acid*** or ***sulfuric acid***. The process of altering these gases into their acid counterparts can take several days, and during this time these pollutants can be transferred hundreds of kilometers from their original source. Acid precipitation formation can also take place at the Earth's surface when nitrogen oxides and sulfur dioxide settle on the landscape and interact with dew or frost.

Emissions of sulfur dioxide are responsible for 60-70 % of the acid deposition that occurs globally. More than 90 % of the sulfur in the atmosphere is of human origin. The main sources of sulfur include:

Coal burning - coal typically contains 2-3 % sulfur so when it is burned sulfur dioxide is liberated.

The smelting of metal sulfide ores to obtain the pure metals. Metals such as zinc, nickel, and copper are all commonly obtained in this manner.

Volcanic eruptions - although this is not a widespread problem, a volcanic eruption can add a lot of sulfur to the atmosphere in a regional area.

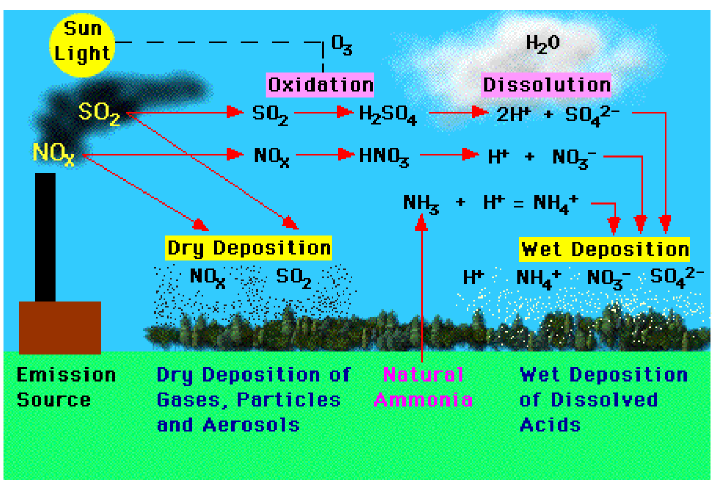
Organic decay.

Ocean spray.

After being released into the atmosphere, sulfur dioxide can either be deposited on the Earth's surface in the form of dry deposition or it can undergo the following reactions to produce acids that are incorporated into the products of wet deposition (**Figure 8h-2**):

$SO_2 + H_2O$ »»» H_2SO_3

$H_2SO_3 + 1/2O_2$ »»» H_2SO_4



{PRIVATE**}Figure 8h-2:** Several processes can result in the formation of acid deposition. ***Nitrogen oxides*** (NO_x) and ***sulfur dioxide*** (SO_2) released into the atmosphere from a variety of sources call fall to the ground simply as dry deposition. This dry deposition can then be converted into acids when these deposited chemicals meet water. Most wet acid deposition forms when nitrogen oxides (NO_x) and sulfur dioxide (SO_2) are converted to ***nitric acid*** (HNO_3) and ***sulfuric acid*** (H_2SO_4) through ***oxidation*** and ***dissolution***. Wet deposition can also form when ***ammonia*** gas (NH_3) from natural sources is converted into ***ammonium*** (NH_4).

Some 95 % of the elevated levels of nitrogen oxides in the atmosphere are the result of human activities. The remaining 5 % comes from several natural processes. The major sources of nitrogen oxides include:

Combustion of oil, coal, and gas.
Bacterial action in soil.
Forest fires.
Volcanic action.
Lightning.

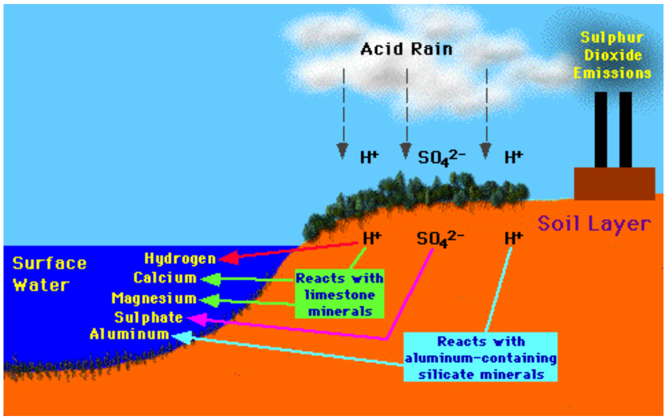
Acids of nitrogen form as a result of the following atmospheric chemical reactions (see **Figure 8h-2** above):

$NO + 1/2O_2$ »»» NO_2
$2NO_2 + H_2O$ »»» $HNO_2 + HNO_3$
$NO_2 + OH$ »»» HNO_3

Finally, the concentrations of both nitrogen oxides and sulfur dioxides are much lower than atmospheric carbon dioxide which is mainly responsible for making natural rainwater slightly acidic. However, these gases are much more soluble than carbon dioxide and therefore have a much greater effect on the pH of the precipitation.

Effects of Acid Deposition

Acid deposition influences the environment in several different ways. In aquatic systems, acid deposition can effect these ecosystems by lowering their pH. However, not all aquatic systems are effected equally. Streams, ponds, or lakes that exist on bedrock or sediments rich in calcium and/or magnesium are naturally buffered from the effects of acid deposition. Aquatic systems on neutral or acidic bedrock are normally very sensitive to acid deposition because they lack basic compounds that buffer acidification (see **Figure 8h-3**). In Canada, many of the water bodies found on the granitic ***Canadian Shield*** fall in this group. One of the most obvious effects of aquatic acidification is the decline in fish numbers. Originally, it was believed that the fish died because of the increasing acidity of the water. However, in the 1970s scientists discovered that acidified lakes also contained high concentrations of toxic heavy metals like mercury, aluminum, and cadmium. The source of these heavy metals was the soil and bedrock surrounding the water body. Normally, these chemicals are found locked in clay particles, minerals, and rocks. However, the acidification of terrestrial soils and bedrock can cause these metals to become soluble. Once soluble, these toxic metals are easily leached by infiltrating water into aquatic systems where they accumulate to toxic levels.



{PRIVATE}**Figure 8h-3:** Lake acidification begins with the deposition of the byproducts acid precipitation (SO4 and H ions) in terrestrial areas located adjacent to the water body. Hydrologic processes then move these chemicals through soil and bedrock where they can react with limestone and aluminum-containing silicate minerals. After these chemical reactions, the ***leachate*** continues to travel until it reaches the lake. The acidity of the leachate entering lake is controlled by the chemical composition of the effected lake's surrounding soil and bedrock. If the soil and bedrock is rich in limestone the acidity of the infiltrate can be reduced by the buffering action of calcium and magnesium compounds. Toxic aluminum (and some other toxic heavy metals) can leach into the lake if the soil and bedrock is rich in aluminum-rich silicate minerals.

In the middle latitudes, many acidified aquatic systems experience a phenomenon known as ***acid shock***. During the winter the acidic deposits can buildup in the snowpack. With the arrival of spring, snowpack begins to melt quickly and the acids are released over a short period of time at concentrations 5 to 10 times more acidic than rainfall. Most adult fish can survive this shock. However, the eggs and small fry of many spring spawning species are extremely sensitive to this acidification.

The severity of the impact of acid deposition on vegetation is greatly dependent on the type of soil the plants grow in. Similar to surface water acidification, many soils have a natural buffering capacity and are able to neutralize acid inputs. In general, soils that have a lot of lime are better at neutralizing acids than those that are made up of siliceous sand or weathered acidic bedrock. In less buffered soils, vegetation is effected by acid deposition because:

Increasing acidity results in the leaching of several important plant nutrients, including calcium, potassium, and magnesium. Reductions in the availability of these nutrients cause a decline in plant growth rates.

The heavy metal aluminum becomes more mobile in acidified soils. Aluminum can damage roots and interfere with plant uptake of other nutrients such as magnesium and potassium.

Reductions in soil pH can cause germination of seeds and the growth of young seedlings to be inhibited.

Many important soil organisms cannot survive is soils below a pH of about 6.0. The death of these organisms can inhibit decomposition and nutrient cycling.

High concentrations of nitric acid can increase the availability of nitrogen and reduce the availability of other nutrients necessary for plant growth. As a result, the plants become over-fertilized by nitrogen (a condition known as ***nitrogen saturation***).

Acid precipitation can cause direct damage to the foliage on plants especially when the precipitation is in the form of fog or cloud water which is up to ten times more acidic than rainfall.

Dry deposition of SO2 and NOx has been found to affect the ability of leaves to retain water when they are under water stress.

Acidic deposition can leach nutrients from the plant tissues weakening their structure.

The combination of these effects can lead to plants that have reduced growth rates, flowering ability and yields. It also makes plants more vulnerable to diseases, insects, droughts and frosts.

The effects of acidic deposition on humans can be divided into three main categories. Acid deposition can influence human health through the following methods:

Toxic metals, such as mercury and aluminum, can be released into the environment through the acidification of soils. The toxic metals can then end up in the drinking water, crops, and fish, and are then ingested by humans through consumption. If ingested in great quantities, these metals can have toxic effects on human health. One metal, aluminum, is believed to be related to the occurrence of **Alzheimer's disease**.

Increased concentrations of sulfur dioxide and oxides of nitrogen have been correlated to increased hospital admissions for respiratory illness.

Research on children from communities that receive a high amount of acidic pollution show increased frequencies of chest colds, allergies, and coughs.

Acid deposition also influences the economic livelihoods of some people. Many lakes and streams on the eastern coast of North America are so acidic that the fish decline significantly in numbers. The reduced fish numbers then influence commercial fishermen and industries that rely on sport fishing tourism. Forestry and agriculture are effected by the damage caused to vegetation. In some areas of eastern North America and Europe, large die-backs of trees have occurred.

Finally, acid deposition effects a number inanimate features of human construction. Buildings and head stones that are constructed from limestone are easily attacked by acids, as are structures that are constructed of iron or steel. Paint on cars can react with acid deposition causing fading. Many of the churches and cathedrals in Europe are under attack from the effects of acidic deposition.

(i) Evaporation and Transpiration

Water is removed from the surface of the Earth to the atmosphere by two distinct mechanisms: ***evaporation*** and ***transpiration***.

Evaporation can be defined as the process where liquid water is transformed into a gaseous state. Evaporation can only occur when water is available. It also requires that the humidity of the atmosphere be less than the evaporating surface (at 100 % ***relative humidity*** there is no more evaporation). The evaporation process requires large amounts of energy. For example, the evaporation of one gram of water requires 600 calories of ***heat energy***.

Transpiration is the process of water loss from plants through ***stomata***. Stomata are small openings found on the underside of leaves that are connected to vascular plant tissues. In most plants, transpiration is a passive process largely controlled by the humidity of the atmospheric and the moisture content of the soil. Of the transpired water passing through a plant only 1 % is used in the growth process. Transpiration also transports nutrients from the soil into the roots and carries them to the various cells of the plant and is used to keep tissues from becoming overheated. Some dry environment plants do have the ability to open and close their stomata. This adaptation is necessary to limit the loss of water from plant tissues. Without this adaptation these plants would not be able to survive under conditions of severe drought.

It is often difficult to detect between evaporation and transpiration. So we use a composite term ***evapotranspiration***. The rate of evapotranspiration at any instant from the Earth's surface is controlled by four factors:

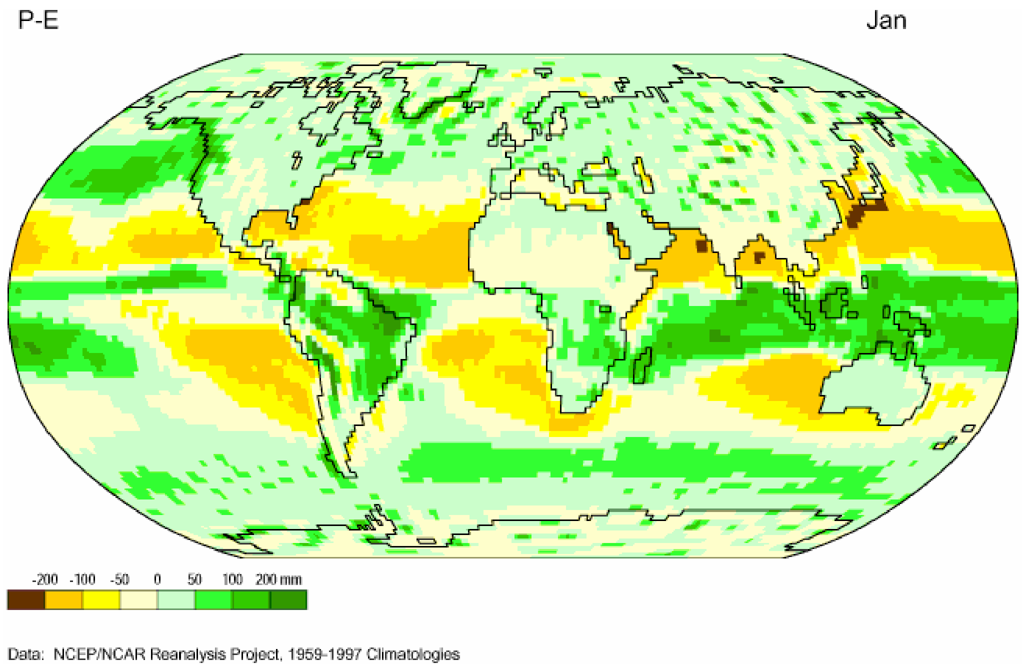
Energy availability. The more energy available the greater the rate of evapotranspiration. It takes about 600 calories of heat energy to change 1 gram of liquid water into a gas.

The humidity gradient away from the surface. The rate and quantity of water vapor entering into the atmosphere both become higher in drier air.

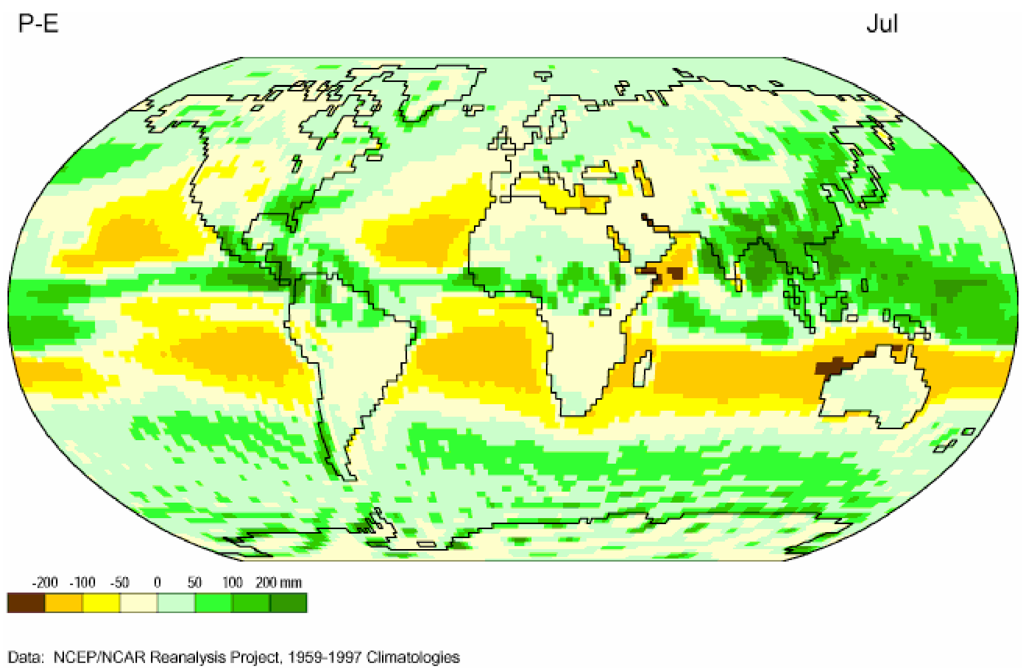
The wind speed immediately above the surface. Many of us have observed that our gardens need more watering on windy days compared to calm days when temperatures are similar. This fact occurs because wind increases the potential for evapotranspiration. The process of evapotranspiration moves water vapor from ground or water surfaces to an adjacent shallow layer that is only a few centimeters thick. When this layer becomes saturated evapotranspiration stops. However, wind can remove this layer replacing it with drier air which increases the potential for evapotranspiration.

Water availability. Evapotranspiration cannot occur if water is not available.

On a global scale, most of the evapotranspiration of water on the Earth's surface occurs in the subtropical oceans (**Figures 8i-1** and **8i-2**). In these areas, high quantities of solar radiation provide the energy required to convert liquid water into a gas. Evapotranspiration generally exceeds precipitation on middle and high latitude landmass areas during the summer season. Once again, the greater availability of solar radiation during this time enhances the evapotranspiration process.



{PRIVATE**}Figure 8i-1:** Precipitation minus evapotranspiration for an average January, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).



{PRIVATE**}Figure 8i-2:** Precipitation minus evapotranspiration for an average July, 1959-1997. (**Source of Original Modified Image:** Climate Lab Section of the Environmental Change Research Group, Department of Geography, University of Oregon - *Global Climate Animations*).

(j) Actual and Potential Evapotranspiration

Often, scientists characterize between two different aspects of evapotranspiration: ***potential evapotranspiration*** and ***actual evapotranspiration***.

Potential evapotranspiration or **PE** is a measure of the ability of the atmosphere to remove water from the surface through the processes of ***evaporation*** and ***transpiration*** assuming no control on water supply. ***Actual evapotranspiration*** or **AE** is the quantity of water that is actually removed from a surface due to the processes of ***evaporation*** and ***transpiration***.

Scientists consider these two types of ***evapotranspiration*** for the practical purpose of water resource management. Around the world humans are involved in the production of a variety of plant crops. Many of these crops grow in environments that are naturally short of water. As a result, irrigation is used to supplement the crop's water needs. Managers of these crops can determine how much supplemental water is needed to achieve maximum productivity by estimating potential and actual evapotranspiration. Estimates of these values are then used in the following equation:

crop water need = potential evapotranspiration - actual evapotranspiration

The following factors are extremely important in estimating potential evapotranspiration:

Potential evapotranspiration requires energy for the evaporation process. The major source of this energy is from the sun. The amount of energy received from the sun accounts for 80 % of the variation in **potential evapotranspiration**.

Wind is the second most important factor influencing **potential evapotranspiration**. Wind enables water molecules to be removed from the ground surface by a process known as ***eddy diffusion***.

The rate of ***evapotranspiration*** is associated to the gradient of ***vapor pressure*** between the ground surface and the layer of atmosphere receiving the evaporated water.

(k) Interception, Stemflow, Canopy Drip, and Throughfall

Vegetation often modifies the intensity and distribution of precipitation falling on and through its leaves and woody structures. The most obvious effect plants have on falling precipitation is ***interception*** (**Figure 8k-1**). **Interception** can be technically defined as the capture of ***precipitation*** by the plant canopy and its subsequent return to the atmosphere through ***evaporation*** or ***sublimation***. The amount of precipitation intercepted by plants varies with leaf type, canopy architecture, wind speed, available radiation, temperature, and the humidity of the atmosphere.

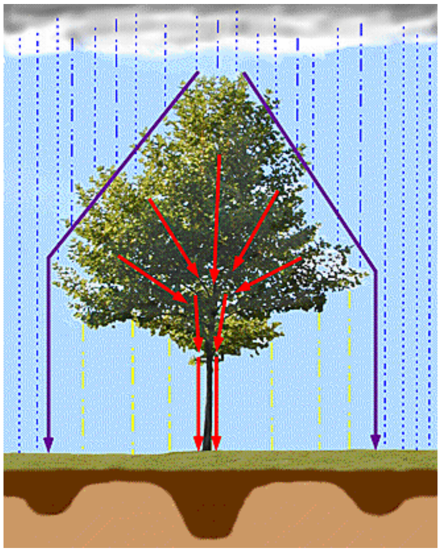
{PRIVATE**}Figure 8k-1:** Vegetation can intercept up to 50 % of the rain that falls on its leaves. The leaves of deciduous trees commonly intercept anywhere from 20 to 30 % of the falling rain. Water dripping off leaves to the ground surface is technically called ***leaf drip***.

Precipitation that is not intercepted can be influenced by the following processes (see **Figure 8k-2** below):

Stemflow - is the process that directs ***precipitation*** down plant branches and stems (red arrows in **Figure 8k-2**). The redirection of water by this process causes the ground area around the plant's stem to receive additional moisture. The amount of stemflow is determined by leaf shape and stem and branch architecture. In general, deciduous trees have more stemflow than ***coniferous*** vegetation.

Canopy drip - some plants have an architecture that directs ***rainfall*** or ***snowfall*** along the edge of the plant canopy (purple arrows in **Figure 8k-1**). This is especially true of ***coniferous*** vegetation. On the ground, canopy drip creates areas with higher moisture content that are located in a narrow band at the edge of the plant canopy.

Throughfall - describes the process of ***precipitation*** passing through the plant canopy (yellow dashed lines in **Figure 8k-2**). This process is controlled by factors like: plant leaf and stem density, type of the precipitation, intensity of the precipitation, and duration of the precipitation event. The amount of precipitation passing through varies greatly with vegetation type.



{PRIVATE**}Figure 8k-2:** Modification of falling precipitation by vegetation The relative quantity of precipitation entering the soil is indicated in dark brown.

(l) Infiltration and Soil Water Storage

Infiltration

Infiltration refers to the movement of water into the ***soil*** layer. The rate of this movement is called the ***infiltration rate***. If rainfall intensity is greater than the infiltration rate, water will accumulate on the surface and ***runoff*** will begin.

Movement of water into the ***soil*** is controlled by ***gravity***, ***capillary action***, and ***soil porosity***. Of these factors soil porosity is most important. A soil's porosity is controlled by its texture, structure, and organic content. Coarse textured soils have larger pores and fissures than fine-grained soils and therefore allow for more water flow. Pores and fissures found in soils can be made larger through a number of factors that enhance internal soil structure. For example, the burrowing of worms and other organisms and penetration of plant roots can increase the size and number of macro and micro-channels within the soil. The amount of decayed ***organic matter*** found at the soil surface can also enhance infiltration. Organic matter is generally more porous than mineral soil particles and can hold much greater quantities of water.

The rate of infiltration normally declines rapidly during the early part of a rainstorm event and reaches a constant value after several hours of rainfall. A number of factors are responsible for this phenomena, including:

(1) The filling of fine soil pores with water reduces capillary forces.

(2) As the soil moistens, clay particles swell and reduce the size of pores.

(3) Raindrop impact breaks up soil clumps, splashing fine particles into pores.

Soil Water Storage

Within the soil system, the storage of water is influenced by several different forces. The strongest force is the molecular force of elements and compounds found on the surface of soil minerals. The water retained by this force is called ***hygroscopic water*** and it consists of the water held within 0.0002 millimeters of the surface of soil particles. The maximum limit of this water around a soil particle is known as the ***hygroscopic coefficient***. Hygroscopic water is essentially non-mobile and can only be removed from the soil through heating. ***Matric force*** holds soil water from 0.0002 to 0.06 millimeters from the surface of soil particles. This force is due to two processes: soil particle surface molecular attraction (adhesion and absorption) to water and the cohesion that water molecules have to each other. This force declines in strength with distance from the soil particle. The force becomes nonexistent past 0.06 millimeters. ***Capillary action*** moves this water from areas where the matric force is low to areas where it is high. Because this water is primarily moved by capillary action, scientists commonly refer to it as ***capillary water***. Plants can use most of this water by way of capillary action until the soil ***wilting point*** is reached. Water in excess of capillary and hygroscopic water is called ***gravitational water***. Gravitational water is found beyond 0.06 millimeters from the surface of soil particles and it moves freely under the effect of gravity. When gravitational water has drained away the amount of water that remains is called the soil's ***field capacity***.

Figure 8l-1 describes the relationship between the thickness of water film around soil particles and the strength of the force that holds this water. Force is measured in units called ***bars***. One bar is equal to a 1000 ***millibars***. The graph also displays the location of ***hygroscopic water***, the ***hygroscopic coefficient***, the ***wilting point***, ***capillary water***, ***field capacity***, and ***gravitational water*** along this line.

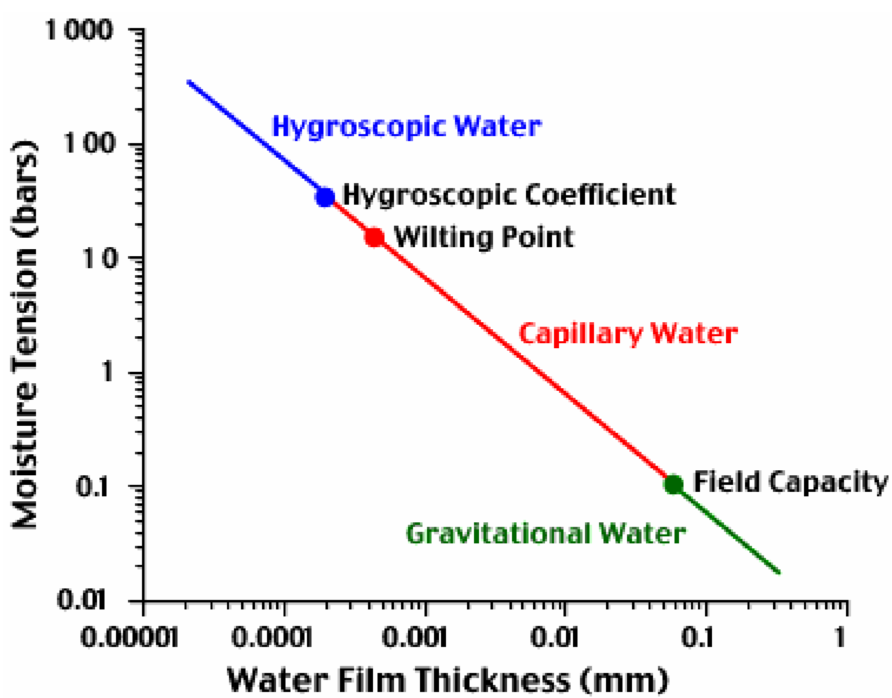
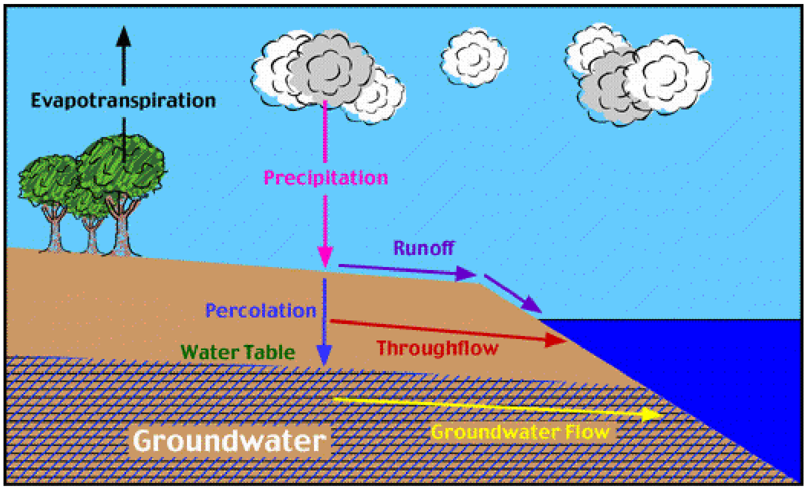


Figure 8l-1: Relationship between soil water film thickness and moisture tension.

(m) Throughflow and Groundwater Storage

Throughflow is the sporadic horizontal flow of water within the soil layer (**Figure 8m-1**). It normally takes place when the soil is completely saturated with water. This water then flows underground until it reaches a ***river***, ***lake***, or ***ocean***. Rates of water movement via throughflow are usually low. Rates of maximum flow occur on steep slopes and in pervious sediments. The lowest rates of flow occur in soils composed of heavy clays. Rates of throughflow in these sediments can be less than 1 millimeter per day.



{PRIVATE}**Figure 8m-1:** Hydrologic movement of water beneath the Earth's surface. Water usually enters the surface sediments as ***precipitation***. This water then ***percolates*** into the soil layer. Some of this water flows horizontally as ***throughflow***. Water continuing to flow downward eventually reaches a permanent store of water known as the ***groundwater***. The movement of groundwater horizontally is called ***groundwater flow***.

Precipitation that succeeds in moving from the soil layer down into the underlying bedrock will at some point reach an area of permanent saturation that is known as the ***groundwater*** zone (**Figure 8m-1**). The top of this zone is called the ***water table***. Approximately 22 % of the fresh water found at the Earth's terrestrial surface is stored as groundwater. Groundwater tends to flow by way of gravity to the point of lowest elevation. Often ***groundwater flow*** discharges into a surface body of water like a ***river channel***, ***lake***, or ***ocean***. Typical groundwater flow velocities lie in the range of 250 to 0.001 meters per day. Highest groundwater flow velocities are commonly found in sedimentary deposits (like ***gravel***, ***conglomerate***, or ***sandstone***) because of their very high ***permeability***. The least permeable ground type occurs in dense ***igneous*** rock materials like ***granite***. Rock formations that store groundwater water are known as ***aquifers***. Rock formations that cannot store groundwater are called ***aquicludes***.

Groundwater occurs in two main forms. ***Unconfined groundwater*** occurs when the flow of subterranean water is not confined by the presence of relatively impermeable layers (**Figure 8m-2**). The presence of an impermeable layer beneath this type of groundwater can cause the formation of a ***perched water table***. These features are elevated some distance above the surface's main water table. ***Springs*** that flow from underground to the Earth's surface are often formed when a perched water table intersects the surface.

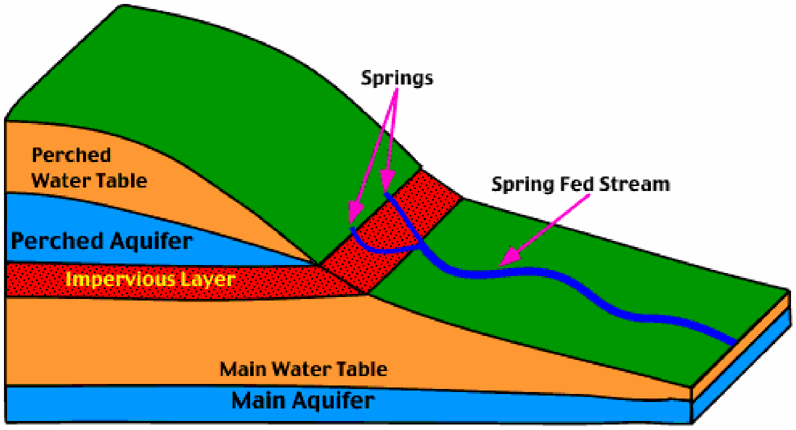


Figure 8m-2: Unconfined groundwater, perched water tables and the development of springs.

In some cases, groundwater can become ***confined*** between two impermeable layers (**Figure 8m-3**). This type of enclosed water is sometimes called ***artesian***. If conditions are right, a confined aquifer can produce a pressurized ground to surface flow of water known as an ***artesian well***. In an artesian well, water flows against gravity to the earth's surface because of ***hydrostatic pressure***. Hydrostatic pressure is created from the fact that most of the aquifer's water resides at an elevation greater than the well opening. The overlying weight of this water creates the hydrostatic pressure.

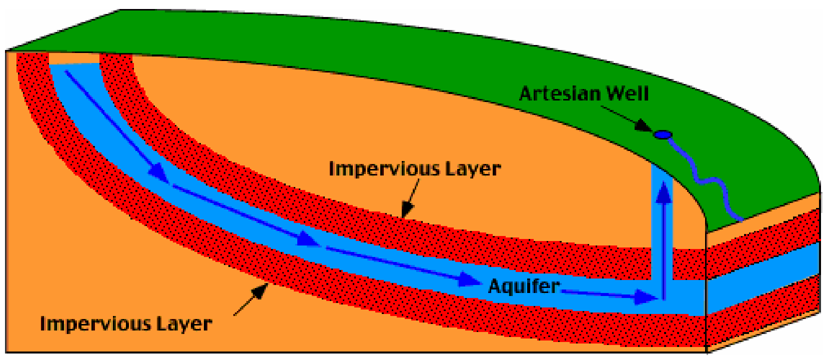


Figure 8m-3: Confined groundwater and the development of an artesian well.

(n) Introduction to Surface Runoff

Surface Runoff

If the amount of water falling on the ground is greater than the ***infiltration rate*** of the surface, ***runoff*** or **overland flow** will occur. **Runoff** specifically refers to the water leaving an area of drainage and flowing across the land surface to points of lower elevation. It is not the water flowing beneath the surface of the ground. This type of water flow is called ***throughflow***. **Runoff** involves the following events:

Rainfall intensity exceeds the soil's ***infiltration rate***.

A thin water layer forms that begins to move because of the influence of slope and gravity.

Flowing water accumulates in depressions.

Depressions overflow and form small ***rills***.

Rills merge to form larger ***streams*** and ***rivers***.

Streams and rivers then flow into ***lakes*** or ***oceans***.

On a global scale, runoff occurs because of the imbalance between ***evaporation*** and ***precipitation*** over the Earth's land and ocean surfaces. Oceans make up 71 % of the Earth's surface and the solar radiation received here powers the global evaporation process. In fact, 86 % of the Earth's evaporation occurs over the oceans, while only 14 % occurs over land. Of the total amount of water evaporated into the atmosphere, precipitation returns only 79 % to the oceans, and 21 % to the land. Surface runoff sends 7 % of the land based precipitation back to the ocean to balance the processes of evaporation and precipitation.

The distribution of **runoff per continent** shows some interesting patterns (see **Table 8n-1**). Areas having the most runoff are those with high rates of precipitation and low rates of evaporation.

{PRIVATE**}Table 8n-1:** Continental runoff values. (**Source**: Lvovitch, M.L. 1972. World water balance, In: **Symposium of World Water Balance**. IASH-UNESCO. Report Number 92).

{PRIVATE**}Continent**	**Runoff Per Unit Area (mm per yr.)**
Europe	300
Asia	286
Africa	139
North and Central America	265
South America	445
Australia, New Zealand and New Guinea	218
Antarctica and Greenland	164

Streamflow and Stream Discharge

The term ***streamflow*** describes the process of water flowing in the organized channels of a ***stream*** or ***river***. ***Stream discharge*** represents the volume of water passing through a river channel during a certain period of time. Stream discharge can be expressed mathematically with the following equation:

$Q = W \times D \times V$

where,

Q equals stream discharge usually measured in cubic meters per second, **W** equals channel width, **D** equals channel depth, and **V** equals velocity of flowing water.

Because of streamflow's potential hazard to humans many streams are gauged by mechanical recorders. These instruments record the stream's discharge on a ***hydrograph***. The **graph** (**Figure 8n-1**) below illustrates a typical hydrograph and its measurement of discharge over time.

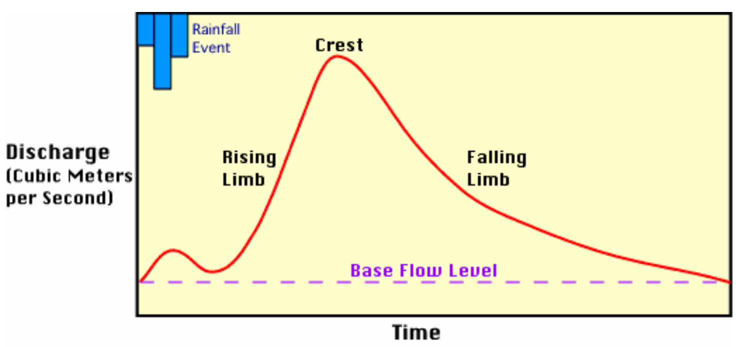


Figure 8n-1: Stream hydrograph.

From this graph we can observe the following things:

A small blip caused by rain falling directly into the channel is the first evidence that stream discharge is changing because of the rainfall.

A significant time interval occurs between the start of rain and the beginning of the main rise in discharge on the hydrograph. This lag occurs because of the time required for the precipitation that falls in the stream's basin to eventually reach the recording station. Usually, the larger the basin the greater the the time lag.

The rapid movement of surface ***runoff*** into the stream's channels and subsequent flow causes the discharge to rise quickly.

The falling limb of the hydrograph tends to be less steep that the rise. This flow represents the water added from distant tributaries and from ***throughflow*** that occurs in surface soils and sediments.

After some time the hydrograph settles at a constant level known as ***base flow*** stage. Most of the base flow comes from ***groundwater flow*** which moves water into the stream channel very slowly.

Not all hydrographs are the same. Actually, the shape and magnitude of the hydrograph is controlled by two sets of factors:

Permanent Factors - slope of basin, soil structure, vegetation, channel density, etc.

Transient Factors - are those factors associated with precipitation input - size of storm, intensity, duration of rainfall, etc.

(o) Introduction to the Oceans

Oceans cover approximately 71 % or 360 million square kilometers of the Earth's surface. On average, the depth of the world's oceans is about 3.9 kilometers. Maximum depths, however, can exceed 11 kilometers! The oceans contain 97 % of our planet's free water. The other 3 % is found in ***atmosphere***, on the Earth's terrestrial surface, or in the Earth's ***lithosphere*** in various forms and stores (see the ***Hydrologic Cycle***).

The distribution of ocean basins and continents is unevenly arranged over the Earth's surface (**Figure 8o-1**). In the Northern Hemisphere, the ratio of land to ocean is about 1 to 1.5. The ratio of land to ocean in the Southern Hemisphere is 1 to 4. The greater abundance of water in the Southern Hemisphere has some interesting effects on the environment of this area. For example, climate tends to be more moderate in the Southern Hemisphere because of the ocean's ability to release large amounts of stored heat energy.

Humans have divided and named the interconnected oceans of the world into three groups: the **Atlantic** (including the **Arctic Sea**), the **Indian**, and the **Pacific.**

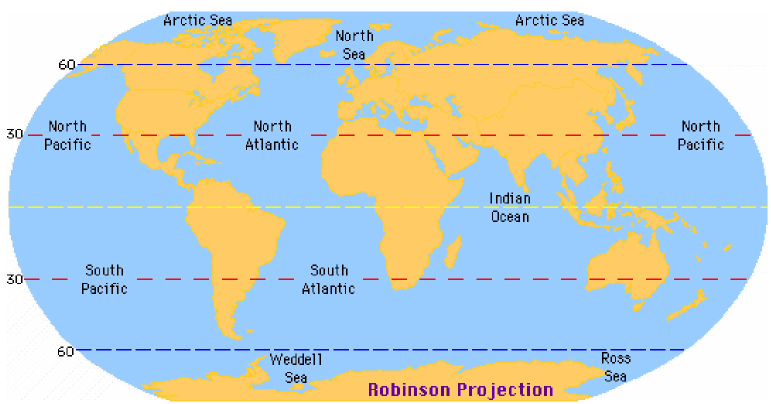


Figure 8o-1: Oceans and seas of the world.

The **Pacific** is the largest ocean basin. It has an average depth of 4.3 kilometers and has few shallow marginal seas, but many islands. Only a few rivers discharge into this ocean basin. This lack of rivers is demonstrated by fact that the surface area of the Pacific is about 1000 percent greater than the land area that drains into it.

The **Atlantic** is a relatively narrow body of water that twists between nearly parallel continental masses. The Atlantic Ocean contains the majority of the Earth's shallow seas, but relatively few islands. Some of the shallow seas found in this basin include: the **Caribbean**, **Mediterranean**, **Baltic**, **Arctic Seas**, and the **Gulf of Mexico**. Many streams and rivers discharge into the Atlantic Ocean. This basin also drains some of the world's largest rivers including the Amazon, Mississippi, St. Lawrence, and Congo. The surface area of the Atlantic Ocean is about 1.6 times greater than the terrestrial area discharging into to it. As a result, the Atlantic Ocean receives more fresh water from continental ***runoff*** than any other ocean basin.

The **Indian** Ocean is the smallest of the three major ocean basins. It is bordered by the landmasses of **Africa** and **Asia**. This basin has few islands and limited shallow seas. The surface area of the Indian ocean is approximately 400 percent larger than the area draining into it. Because of its close proximity to the equator this basin has the warmest surface ocean temperatures.

(p) Physical and Chemical Characteristics of seawater

Seawater is a mixture of various salts and water. Most of the water in the ocean basins is believed to originate from the ***condensation*** of water found in the early atmosphere as the Earth cooled after its formation. This water was released from the ***lithosphere*** as the Earth's crust solidified. Additional water has also been added to the oceans over geologic time from periodic ***volcanic*** action. Some scientists have recently speculated that comets entering the Earth's atmosphere may be another important source of water for the oceans.

Most of the dissolved chemical constituents or salts found in seawater have a continental origin. It seems that these chemicals were released from continental rocks through ***weathering*** and then carried to the oceans by stream ***runoff***. Over time, the concentration of these chemicals increased until an equilibrium was met. This equilibrium occurred when the ocean's water could not dissolve any more material in solution. Similarities between fossilized sea life and organisms living today indicate that the composition of seawater stopped changing drastically about 600 million years ago.

Only six ***elements*** comprise about 99 % of sea salts: chlorine (Cl^-), sodium (Na^-), sulfur (SO_4^{-2}), magnesium (Mg^{+2}), calcium (Ca^{-2}), and potassium (K^+) (**Figure 8p-1**). The relative abundance of the major salts in seawater are constant regardless of the ocean. Only the amount of water in the mixture varies because of differences between ocean basins because of regional differences in freshwater loss (evaporation) and gain (runoff and precipitation).

The chlorine ion makes up 55 % of seawater. Calculations of seawater salinity are made of the parts per 1000 of the chlorine ion present in one kilogram of seawater. Typically, seawater has a salinity of 35 parts per thousand.

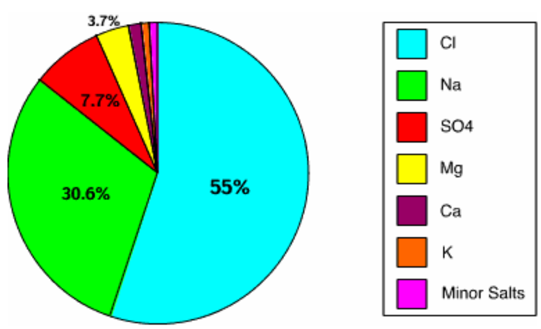


Figure 8p-1: Relative proportions of dissolved salts in seawater.

Water is one of the few substances existing on the Earth's surface in all three forms of matter. At zero degrees Celsius liquid water turns into ice and has a density of approximately 917 kilograms per cubic meter. Liquid water at the same temperature has a density of nearly 1,000 kilograms per cubic meter. The density of seawater generally increases with decreasing temperature, increasing salinity, and increasing depth in the ocean. The density of seawater at the surface of the ocean varies from 1,020 to 1,029 kilograms per cubic meter. Highest densities are achieved with depth because of the overlying weight of water. In the deepest parts of the oceans, seawater densities can be as high as 1,050 kilograms per cubic meter.

Seawater freezes at a temperature that is slightly colder than fresh water (0.0 degrees Celsius). The freezing temperature of seawater also varies with the concentration of salts. More salt the lower the initial freezing temperature. At a salinity of 35 parts per thousand, seawater freezes at a temperature of -1.9 degrees Celsius.

Sea ice normally contains considerably less salt than seawater. Most of the salts found in liquid seawater are forced out it when freezing occurs. The reason for the exclusion is because the molecules of the various salts do not fit well in the highly orderly molecular structure of frozen water. Because of the density difference between ice and seawater, ice floats on the surface of the ocean.

Seawater also contains small amounts of dissolved gases. Many of these gases are added to seawater from the atmosphere through the constant stirring of the sea surface by wind and waves. The concentration of gases that can be dissolved into seawater from the atmosphere is determined by temperature and salinity of the water. Increasing the temperature or salinity reduces the amount of gas that ocean water can dissolve. Some of the important atmospheric gases found in seawater include: nitrogen, oxygen, carbon dioxide (in the form of bicarbonate HCO_3), argon, helium, and neon. Compared to the other atmospheric gases, the amount of ***carbon dioxide*** dissolved in saturated seawater is unusually large.

Some gases found within seawater are also involved in oceanic ***organic*** and ***inorganic*** processes that are indirectly related to the atmosphere. For example, oxygen and carbon dioxide may be temporally generated or depleted by such processes to varying concentrations at specific locations within the ocean.

(q) Surface and Subsurface Ocean Currents

Surface Ocean Currents

An ***ocean current*** can be defined as a horizontal movement of seawater at the ocean's surface. Ocean currents are driven by the circulation of wind above surface waters. Frictional stress at the interface between the ocean and the wind causes the water to move in the direction of the wind. Large ocean currents are a response of the atmosphere and ocean to the flow of energy from the tropics to polar regions. In some cases, currents are transient features and affect only a small area. Other ocean currents are essentially permanent and extend over large horizontal distances.

On a global scale, large ocean currents are constrained by the continental masses found bordering the three oceanic basins. Continental borders cause these currents to develop an almost closed circular pattern called a ***gyre***. Each ocean basin has a large gyre located at approximately 30 degrees North and South latitude in the subtropical regions. The currents in these gyres are driven by the atmospheric flow produced by the subtropical high pressure systems. Smaller gyres occur in the North Atlantic and Pacific Oceans centered at 50 degrees North. Currents in these systems are propelled by the circulation produced by polar low pressure centers. In the Southern Hemisphere, these gyre systems do not develop because of the lack of constraining land masses.

A typical gyre displays four types of joined currents: two east-west aligned currents found respectively at the top and bottom ends of the gyre; and two boundary currents oriented north-south and flowing parallel to the continental margins. Direction of flow within these currents is determined by the direction of the macro-scale wind circulation. Boundary currents play a role in redistributing global heat latitudinally.

Surface Currents of the Subtropical Gyres

On either side of the equator, in all ocean basins, there are two west flowing currents: the **North** and **South Equatorial** (***Figure 8q-1***)**.** These currents flow between 3 and 6 kilometers per day and usually extend 100 to 200 meters in depth below the ocean surface. The **Equatorial Counter Current**, which flows towards the east, is a partial return of water carried westward by the North and South Equatorial currents. In ***El Nino*** years, this current intensifies in the Pacific Ocean.

Flowing from the equator to high latitudes are the **western boundary currents**. These warm water currents have specific names associated with their location: North Atlantic - Gulf Stream; North Pacific - Kuroshio; South Atlantic - Brazil; South Pacific - East Australia; and Indian Ocean - Agulhas. All of these currents are generally narrow, jet like flows that travel at speeds between 40 and 120 kilometers per day. Western boundary currents are the deepest ocean surface flows, usually extending 1000 meters below the ocean surface.

Flowing from high latitudes to the equator are the **eastern boundary currents**. These cold water currents also have specific names associated with their location: North Atlantic - Canary; North Pacific - California; South Atlantic - Benguela; South Pacific - Peru; and Indian Ocean - West Australia. All of these currents are generally broad, shallow moving flows that travel at speeds between 3 and 7 kilometers per day.

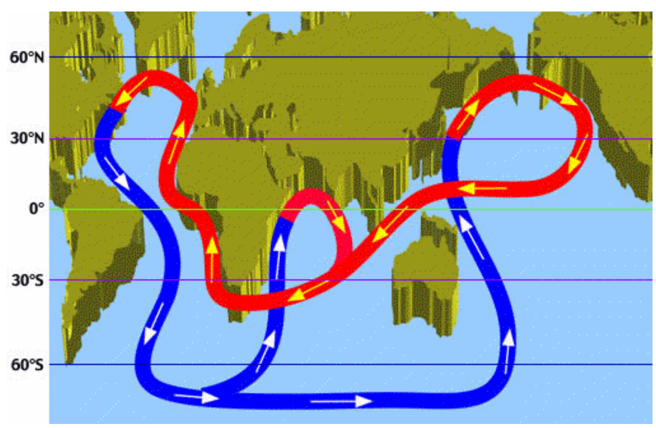
In the Northern Hemisphere, the east flowing **North Pacific Current** and **North Atlantic Drift** move the waters of western boundary currents to the starting points of the eastern boundary currents. The **South Pacific Current**, **South Indian Current** and **South Atlantic Current** provide the same function in the Southern Hemisphere. These currents are associated with the **Antarctic Circumpolar** (**West Wind Drift**). Because of the absence of landmass at this latitude zone, the **Antarctic Circumpolar** flows in continuous fashion around Antarctica and only provides a partial return of water to the three Southern Hemispheric ocean basins.

Surface Currents of the Polar Gyres

The polar gyres exist only in the Atlantic and Pacific basins in Northern Hemisphere. They are propelled by the counterclockwise winds associated with the development of permanent low pressure centers at 50 degrees of latitude over the ocean basins. Note that the bottom west flowing current of the polar gyres is the topmost flowing current of the subtropical gyres. Other currents associated with these gyres are shown on ***Figure 8q-1***.

Subsurface Currents

The world's oceans also have significant currents that flow beneath the surface (**Figure 8q-2**). Subsurface currents generally travel at a much slower speed when compared to surface flows. The subsurface currents are driven by differences in the density of seawater. The density of seawater deviates in the oceans because of variations in temperature and salinity. Near surface seawater begins its travel deep into the ocean in the North Atlantic. The downwelling of this water is caused by high levels of evaporation which cools and increases the salinity of the seawater located here. This seawater then moves south along the coast of North and South America until it reaches Antarctica. At Antarctica, the cold and dense seawater then travels eastward. During this part of its voyage the flow splits off into two currents that move northward. In the North Pacific (off the coast of Asia) and in the Indian Ocean (off the coast of Africa), these two currents move from the ocean floor to its surface creating upwellings. The flow then becomes near surface moving back to the starting point in the North Atlantic. One complete circuit of this flow of seawater is estimated to take about 1,000 years.



{PRIVATE**}Figure 8q-2:** The following illustration describes the flow pattern of the major subsurface ocean currents. Near surface warm currents are drawn in **red**. **Blue** depicts the deep cold currents. Note how this system is continuously moving water from the surface to deep within the oceans and back to the top of the ocean.

(r) Ocean Tides

Introduction

An ocean ***tide*** refers to the cyclic rise and fall of seawater. Tides are caused by slight variations in ***gravitational*** attraction between the **Earth** and the **moon** and the ***sun*** in geometric relationship with locations on the Earth's surface. Tides are periodic primarily because of the cyclical influence of the ***Earth's rotation***.

The moon is the primary factor controlling the temporal rhythm and height of tides (**Figure 8r-1**). The moon produces two tidal bulges somewhere on the Earth through the effects of gravitational attraction. The height of these tidal bulges is controlled by the moon's gravitational force and the Earth's gravity pulling the water back toward the Earth. At the location on the Earth closest to the moon, seawater is drawn toward the moon because of the greater strength of gravitational attraction. On the opposite side of the Earth, another tidal bulge is produced away from the moon. However, this bulge is due to the fact that at this point on the Earth the force of the moon's gravity is at its weakest. Considering this information, any given point on the Earth's surface should experience two tidal crests and two tidal troughs during each tidal period.

{PRIVATE**}Figure 8r-1:** The moon's gravitational pull is the primary force responsible for the tides on the Earth. Photo taken by the Galileo spacecraft from a distance of about 6.2 million kilometers from Earth, on December 16, 1992. (**Source:** *NASA*).

The timing of tidal events is related to the Earth's rotation and the revolution of the moon around the Earth. If the moon was stationary in space, the tidal cycle would be 24 hours long. However, the moon is in motion revolving around the Earth. One revolution takes about 27 days and adds about 50 minutes to the tidal cycle. As a result, the ***tidal period*** is 24 hours and 50 minutes in length.

The second factor controlling tides on the Earth's surface is the sun's gravity. The height of the average solar tide is about 50 % the average lunar tide. At certain times during the moon's revolution around the Earth, the direction of its gravitational attraction is aligned with the sun's (**Figure 8r-2**). During these times the two tide producing bodies act together to create the highest and lowest tides of the year. These ***spring tides*** occur every 14-15 days during full and new moons.

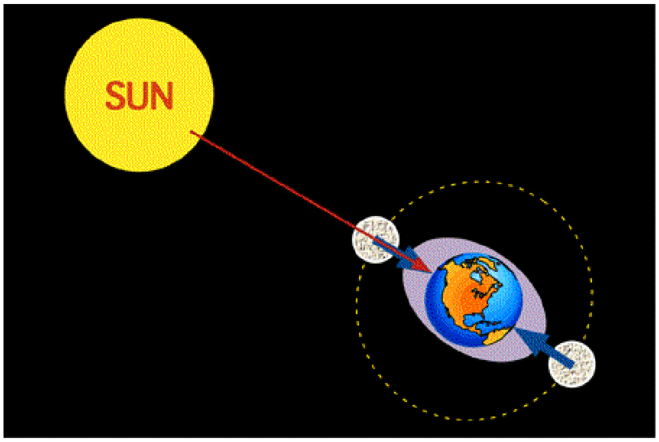


Figure 8r-2: Forces involved in the formation of a spring tide.

When the gravitational pull of the moon and sun are at right angles to each other, the daily tidal variations on the Earth are at their least (**Figure 8r-3**). These events are called ***neap tides*** and they occur during the first and last quarter of the moon.

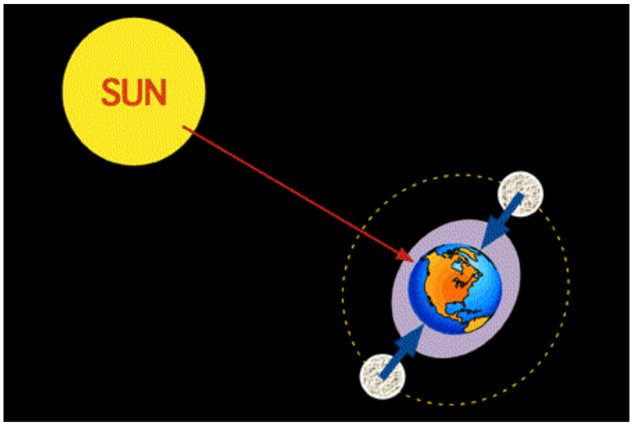


Figure 8r-3: Forces involved in the formation of a neap tide.

Types of Tides

The geometric relationship of moon and sun to locations on the Earth's surface results in creation of three different types of tides. In parts of the northern Gulf of Mexico and Southeast Asia, tides have one high and one low water per tidal day (**Figure 8r-4**). These tides are called ***diurnal tides***.

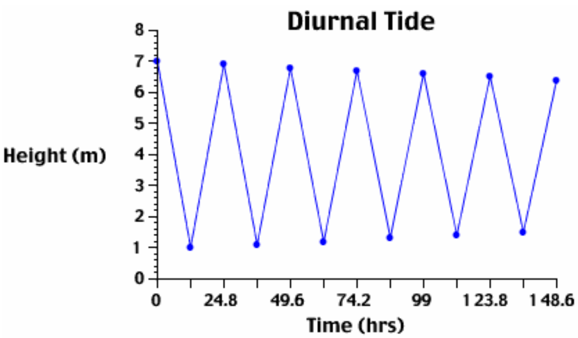


Figure 8r-4: Cyclical tidal cycles associated with a diurnal tide.

Semi-diurnal tides have two high and two low waters per tidal day (**Figure 8r-5**). They are common on the Atlantic coasts of the United States and Europe.

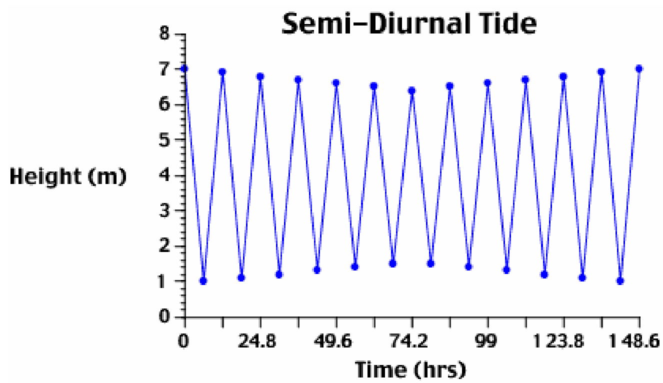


Figure 8r-5: Cyclical tidal cycles associated with a semi-diurnal tide.

Many parts of the world experience ***mixed tides*** where successive high-water and low-water stands differ appreciably (**Figure 8r-6**). In these tides, we have a higher high water and lower high water as well as higher low water and lower low water. The tides around west coast of Canada and the United States are of this type.

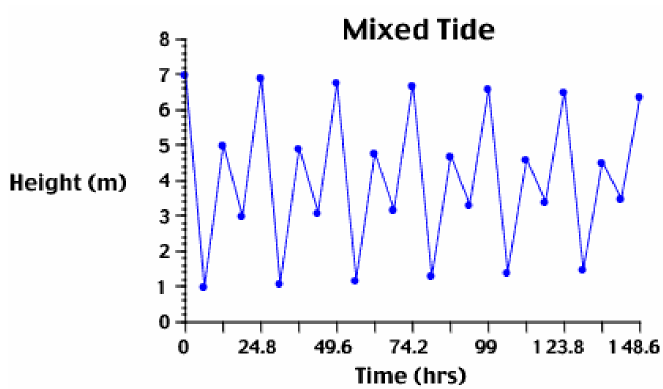


Figure 8r-6: Cyclical tidal cycles associated with a mixed tide.

9) Introduction to Biogeography and Ecology

(a) Origin and Definition of Life

Origin of Life

The sun and its planets formed between 5 and 4.6 billion years ago as matter in our solar system began to coalesce because of ***gravity***. By about 3.9 billion years ago, the Earth had an atmosphere that contained the right mix of hydrogen, oxygen, carbon, and nitrogen to allow for the creation of life. Scientists believe that the energy from heat, lightning, or radioactive elements caused the formation of complex ***proteins*** and ***nucleic acids*** into strands of replicating genetic code. These molecules then organized and evolved to form the first simple forms of life. At 3.8 billion years ago, conditions became right for the fossilization of the Earth's early ***cellular*** life forms. These fossilized ***cells*** resemble present day ***cyanobacteria***. Such cells are known as ***prokaryotes***. Prokaryote cells are very simple, containing few specialized cellular structures and their ***DNA*** is not surrounded by a membranous envelope. The more complex cells of animals and plants, known as ***eukaryotes***, first showed up about 2.1 billion year ago. Eukaryotes have a membrane-bound nucleus and many specialized structures located within their cell boundary. By 680 million years ago, eukaryotic cells were beginning to organize themselves into multicellular organisms. Starting at about 570 million years ago an enormous diversification of multicellular life occurred known as the ***Cambrian explosion***. During this period all but one modern ***phylum*** of animal life made its first appearance on the Earth. **Table 9a-1** describes the approximate time of origin of the Earth' s major groups of plants and animals.

Table 9a-1: Time of origin of some major plant and animal groups.

{PRIVATE**}Organism Group**	**Time of Origin**
Marine Invertebrates	570 Million Years Ago
Fish	505 Million Years Ago
Land Plants	438 Million Years Ago
Amphibians	408 Million Years Ago
Reptiles	320 Million Years Ago
Mammals	208 Million Years Ago
Flowering Plants (***Angiosperms***)	140 Million Years Ago
Hominid Line Begins Its Evolution	20-15 Million Years Ago

Definition of Life

Scientists currently recognize four groups of biological entities:

Archaea - are a group of recently discovered organisms that sometimes live in extremely hostile habitats like thermal volcanic vents, saline pools, and hot springs. Archaea are single-celled organisms that are similar in appearance to ***bacteria***. However, they are biochemically and genetically very different from bacteria. Many books and other forms of scientific literature refer to them as ***archaebacteria***.

Bacteria - are simple single-celled organisms that generally lack ***chlorophyll*** (an exception is ***cyanobacteria*** (see **WWW *Link***). Bacteria have a ***prokaryote*** cell type. They also generally obtain energy for survival through the breakdown of organic matter via ***fermentation*** and ***respiration***. Bacteria such as Rhizobium spp. and ***cyanobacteria*** (see **WWW *Link***) play an important role in the ***fixing*** of atmospheric nitrogen. Without these bacteria ecosystems would be severely short of nitrogen for plant and animal growth. The oldest fossils of life on Earth are bacteria-like organisms.

Eukaryota - are all organisms with a ***eukaryote*** cell type. This group of life includes the ***kingdoms Protista***, ***Fungi***, ***Animalia***, and ***Plantae***.

Viruses - are fragments of ***DNA*** or ***RNA*** that depend on host cells that they infect for their reproduction. They are not cells. Viruses are thought to be parts of the genetic code that originated from either ***eukaryote*** or ***prokaryote*** cells. These code fragments contain enough genetic information for self-existence. At times, viruses are metabolically inert and technically non-living. Viruses cause a variety of diseases in eukaryote organisms. In humans they can cause smallpox, chicken pox, influenza, shingles, herpes, polio, ebola, AIDS, rabies, and some types of cancer.

The four main types of biological entities described above share some unique characteristics that can allow us to distinguish them from non-living things. These characteristics are:

(1). Organisms tend to be complex and highly organized. Chemicals found within their bodies are synthesized through metabolic processes into structures that have defined purposes. ***Cells*** and their various ***organelles*** are examples of such structures. Cells are also the basic functioning unit of life. Cells are often organized into ***organs*** to create higher levels of complexity and function.

(2). Living things have the ability to take energy from their environment and change it from one form to another. This energy is usually used to facilitate their growth and reproduction. We call the process that allows for this facilitation ***metabolism***.

(3). Organisms tend to be ***homeostatic***. In other words, they regulate their bodies and other internal structures to certain normal parameters.

(4). Living creatures respond to stimuli. Cues in their environment cause them to react through behavior, metabolism, and physiological change.

(5). Living things reproduce themselves by making copies of themselves. Reproduction can either be ***sexual*** or ***asexual***. Sexual reproduction involves the fusing of ***haploid*** genetic material from two individuals. This process creates populations with much greater ***genetic diversity***.

(6). Organisms tend to grow and develop. Growth involves the conversion of consumed materials into biomass, new individuals, and waste.

(7). Life adapts and evolves in step with external changes in the environment through ***mutation*** and ***natural selection***. This process acts over relatively long periods of time.

(b) Biological Classification of Organisms

{PRIVATE**}**The term ***species*** has its origins in the ancient Latin language. In this language, the word species means kind. A more technical definition of species is a group of interbreeding organisms that do not ordinarily breed with members of other groups. Biologists estimate that about 10 to 40 million different species inhabit the Earth. Of these species, approximately 1.5 million have been classified.

The first individual to propose an orderly system for classifying the variety of organisms found on our planet was ***Linnaeus*** (1753). In his system of classification, the finest unit in the organization of life is the species. **Linnaeus** suggested that every organism should be classified with a unique binomial name. The first term in this classification system is the organism's generic name or ***Genus***. The second term is the organism's specific name or ***species*** designation.

Current classification systems have developed from **Linnaeus**' original work. However, modern classification systems are much more complicated having many levels of hierarchical organization. These systems are also ***taxonomic*** (structural and physiological connections between organisms), ***phylogenic*** (classification based on genetic connections between organisms), and are structurally based on Darwin's theory of ***evolution***.

Modern classifications of organisms are standardized in a hierarchical system that go from general to specific. **Table 9b-1** below describes the detailed classification of the tree **red maple**. Note that each level of organization is based on some biological characteristic that the organism possesses.

Table 9b-1: Hierarchical system of the biological classification of an organism.

{PRIVATE**}Category**	**Name**	**Characteristics**
Kingdom	Plantae	Organisms that usually have rigid cell walls and usually possess chlorophyll.
Subkingdom	Embryophyta	Plants forming embryos.
Phylum	Tracheophyta	Vascular plants.
Subphylum	Pterophytina	Generally large, conspicuous leaves, complex vascular system.
Class	Angiospermae	Flowering plants, seed enclosed in ovary.
Subclass	Dicotyledoneae	Embryo with two seed leaves.
Order	Sapindales	Soapberry order consisting of a number of trees and shrubs.
Family	Aceraceae	Maple family.
Genus	Acer	Maples and box elder.
Species	Acer rubrum	Red maple.

(c) Natural Selection and Evolution

{PRIVATE}In 1859, **Charles Darwin** published the book ***The Origin of Species*** that presented a theory to explain how life's diversity came to be. It has been suggested that without this theory nothing would make much sense in the field of modern biology.

Evolution describes the process by which species come to possess genetic ***adaptations*** to their environment. Its mechanism is ***natural selection***. Natural selection acts through individuals by determining which individuals have the best adaptations to guarantee reproductive success. Because the state of environment is always changing temporally, natural selection is always influencing the genetic characteristics of the population over time. Thus, natural selection acts to adapt the population to its ever changing surrounds.

Evolutionary change is also a change in ***gene frequency***. The pattern of genetic variation changes from one generation to the next as natural selection determines which individuals are fittest. New ***genes*** enter the species ***gene pool*** by way of ***mutations***. By selecting those organisms that will become reproductively successful, natural selection controls the future frequency of a population's genes. The appearance of new mutations in a population together with the change in gene frequency results in evolution.

The ***spatial isolation*** of sub-populations from a main population is also an important condition for evolution. If a remote sub-population cannot trade genes to the main population because of isolation, natural selection acts differently in each population resulting in ***divergent evolution***. Differences in gene frequency emerge between the two populations because no two patches of habitat are absolutely identical, and because new adaptations enter the gene pool through mutations at the level of the individual. Isolation can be due to distance, mountain barriers, a river, etc.

The best known example of natural selection operating in a modern species is the development of pesticide resistance in many insect species. Prior to the extensive use of pesticides that began in the 1940s, crop pest insect species only contained a small amount of genetic variability for resistance to these chemicals. Natural selection in the absence of pesticides could not lead to changes in the frequencies of genes causing resistance to chemical pesticides. However, once the spraying of pesticides started, individuals that happened to possess resistant genes became much more frequent as they survived the applications. Further, they were able to pass a greater percentage of their genes on to the next generation's gene pool. As a result, the introduction of pesticides into the environment provided a tremendous selective pressure to increase the frequency of resistant genes in the pest populations.

(d) Organization of Life: Species, Populations, Communities, and Ecosystems

{PRIVATE}Scientists have recognized that life can be organized into several different levels of function and complexity. These functional levels are: **species**, **populations**, **communities**, and **ecosystems**.

Species

Species are the different kinds of organisms found on the Earth. A more exact definition of species is **a group of interbreeding organisms that do not ordinarily breed with members of other groups**. If a species interbreeds freely with other species, it would no longer be a distinctive kind of organism. This definition works well with animals. However, in some plant species fertile crossings can take place among morphologically and physiologically different kinds of vegetation. In this situation, the definition of species given here is not appropriate.

Populations

A ***population*** comprises all the individuals of a given species in a specific area or region at a certain time. Its significance is more than that of a number of individuals because not all individuals are identical. Populations contain **genetic variation** within themselves and between other populations. Even fundamental genetic characteristics such as hair color or size may differ slightly from individual to individual. More importantly, not all members of the population are equal in their ability to survive and reproduce.

Communities

Community refers to all the populations in a specific area or region at a certain time. Its structure involves many types of ***interactions*** among species. Some of these involve the acquisition and use of food, space, or other environmental resources. Others involve nutrient cycling through all members of the community and mutual regulation of population sizes. In all of these cases, the structured interactions of populations lead to situations in which individuals are thrown into life or death struggles.

In general, ecologists believe that a ***community*** that has a high ***diversity*** is more **complex** and **stable** than a community that has a low diversity. This theory is founded on the observation that the ***food webs*** of communities of high diversity are more interconnected. Greater interconnectivity causes these systems to be more **resilient** to ***disturbance***. If a species is removed, those species that relied on it for food have the option to switch to many other species that occupy a similar role in that ecosystem. In a low diversity ecosystem, possible substitutes for food may be non-existent or limited in abundance.

Ecosystems

Ecosystems are dynamic entities composed of the biological **community** and the ***abiotic*** environment. An ecosystem's abiotic and biotic composition and structure is determined by the state of a number of interrelated environmental factors. Changes in any of these factors (for example: nutrient availability, temperature, light intensity, grazing intensity, and species population density) will result in dynamic changes to the nature of these systems. For example, a fire in the temperate deciduous forest completely changes the structure of that system. There are no longer any large trees, most of the mosses, herbs, and shrubs that occupy the forest floor are gone, and the nutrients that were stored in the biomass are quickly released into the soil, atmosphere and hydrologic system. After a short time of recovery, the community that was once large mature trees now becomes a community of grasses, herbaceous species, and tree seedlings.

(e) Abiotic Factors and the Distribution of Species

{PRIVATE**}Introduction**

The geographic distributions of plant and animal ***species*** are never fixed over time. ***Geographic ranges*** of organisms shift, expand, and contract. These changes are the result of two contrasting processes: ***colonization*** and ***establishment*** and localized ***extinction***. Colonization and establishment takes place when ***populations*** expand into new areas. A number of processes can initiate this process including ***disturbance*** and ***abiotic*** environmental change. Localized extinction results in the elimination of populations from all or part of their former range. It can be caused by ***biotic interactions*** or, once again, abiotic environmental change.

Dispersal, Colonization, and the Establishment of Species

Many of the different types of organisms that inhabit the Earth have the ability to move. This movement can be accomplished either passively or actively. Active movement requires the organism to use some appendage to initiate walking, running, flying or swimming. In passive movement, the organism uses some external force to cause transit. For example, many plants use wind for ***seed dispersal***, while oyster larvae can passively travel great distances by sea currents.

One common reason why organisms move is to disperse to new habitats. ***Dispersal*** can be defined as the movement of individuals away from others of the same species. In most cases, organisms disperse to escape the influence of their parents and siblings. However, dispersal may also involve a large element of discovery. By finding new suitable habitats individuals can increase the range of their species. A larger range makes the species better off in terms of ***evolution***.

Plants have developed a number of different mechanisms for dispersing their offspring. Some of the common techniques include:

The use of specialized morphological structures to aid the transport of an individual by wind;

The employment of particular morphological structures to transport the individual by moving water;

The production of fruit encased seeds that other organisms consume and disperse;

Adhesion mechanisms; and

The physical ejection of seeds.

Once dispersed, an individual can only ***colonize*** a new site if it is devoid of other organisms and if the necessary ***abiotic*** requirements and conditions exist for its survival. Sites within ecosystems become devoid of organisms through ***disturbance***. A disturbance can be caused by predation, climate variations, earthquakes, volcanoes, fire, animal burrowing, and even the impact of a raindrop.

Often the struggle for survival does not end with colonization of an individual on a vacant site. Once colonized an individual may not be able to ***establish*** itself over the longterm because of abiotic and biotic influences. The death of the individual may occur through competitive interaction, predation or an abiotic factor like fire.

Abiotic Factors and Tolerance Limits

Most species appear to be limited in at least part of their geographic range by ***abiotic*** factors, such as temperature, moisture availability, and soil nutrients. No species is adapted to survive under all conditions found on the Earth. All species have specific limits of tolerance to physical factors that directly effect their survival or reproductive success. The portion of the abiotic factor's range of variation which a species can survive and function in is commonly defined as the ***tolerance range***. The level within the tolerance range at which a species or population can function most efficiently is termed the ***optimum***.

In 1840, **J. Liebig** suggested that organisms are generally limited by only one single physical factor that is in shortest supply relative to demand. At one time ecologists accepted this idea so completely that they called it **Liebig's *Law of the Minimum*** and tried to determine the single ***limiting factor*** that controls the growth of numerous species. However, many studies have shown that Liebig's concept is inadequate to account for the distributional limits of a large number of species. In these cases, scientists believe that complex interactions between several physical factors are responsible for distribution patterns.

(f) Biotic Interactions and the Distribution of Species

{PRIVATE**}Introduction**

Interacting species have a tremendous influence on the size of each other's ***populations***. The various mechanisms for these ***biotic*** influences are quite different from the way in which ***abiotic*** factors effect the size of populations. ***Biotic*** factors also regulate the size of populations more intensely. Finally, the influence of biotic interactions can occur at two different levels. ***Interspecific*** effects are direct interactions between species, and the ***intraspecific*** effects represent interactions of individuals within a single species.

Neutralism

Neutralism is the most common type of interspecific interaction. Neither ***population*** directly affects the other. What interactions occur are slight and indirect. The simple presence of the two ***species*** should not directly affect the population level of either. An example of neutralism would be the interaction between rainbow trout and dandelions living in a mountain valley.

Competition

When two or more organisms in the same community seek the same ***resource*** (e.g., food, water, nesting space, ground space), which is in limiting supply to the individuals seeking it, they ***compete*** with one another. If the ***competition*** is among members of the same ***species***, it is called **intraspecific**. Competition among individuals of different species it is referred to as **interspecific competition**. Individuals in ***populations*** experience both types of competition to a greater or lesser degree.

Competition may be the result of two different processes: ***exploitation*** or ***interference***. Competition by ***exploitation*** occurs between individuals when the indirect effects of two or more species or individuals reduce the supply of the limiting ***resource*** or resources needed for survival. The exclusion of one organism by another can only occur when the dominant organism requires less of the limiting resource to survive. Further, the dominant species must be able to reduce the quantity of the resource to some critical level with respect to the other organism. Resource exploitation, however, does not always cause the exclusion of a species from a community. It may just cause the species involved in this interaction to experience a reduction in their potential growth.

Competition by ***interference*** occurs when an individual directly prevents the physical establishment of another individual in a portion of a habitat. Established plants can preempt the invasion and colonization of other individuals by way of dense root mats, peat and litter accumulation, and mechanical abrasion.

Amensalism

Amensalism is an interaction where one ***species*** suffers and the other interacting species experiences no effect. One particular form of amensalism is ***allelopathy*** which occurs with plants. Allelopathy involves the production and release of chemical substances by one species that inhibit the growth of another. Allelopathic substances range from acids to bases to simple ***organic*** compounds. All of these substances are known under the general term: ***secondary substances***. Secondary substances are chemicals produced by plants that seem to have no direct use in ***metabolism***. A good example of a secondary substance is the antibiotic juglone which is secreted by Black Walnut (Juglans nigra) trees. This substance is known to inhibit the growth of trees, shrubs, grasses, and herbs found growing near Black Walnut trees. In the chaparral vegetation of California, certain species of shrubs, notably Salvia leucophylla (mint) and Artemisia californica (sagebrush) are known to produce allelopathic substances. Often these chemicals accumulate in the soil during the dry season reducing the germination and growth of grasses and herbs in an area up to 1 to 2 meters from the secreting plants.

Mutualism

Mutualism is the name given to associations between pairs of ***species*** that bring mutual benefit. The individuals in the ***populations*** of each mutualist species grow and/or survive and/or reproduce at a higher rate when in the presence of individuals of the other species. In most ecology or biogeography textbooks mutualisms are generally underemphasized or ignored. Yet this type of interaction is an extremely widespread phenomena. For example, most rooting plants have mutualistic associations with fungal ***mycorrhizae***. Mycorrhizae increase the capability of plant roots to absorb ***nutrients*** like nitrogen and phosphorus. In return, the roots of the host provide support and a constant supply of ***carbohydrates*** for consumption.

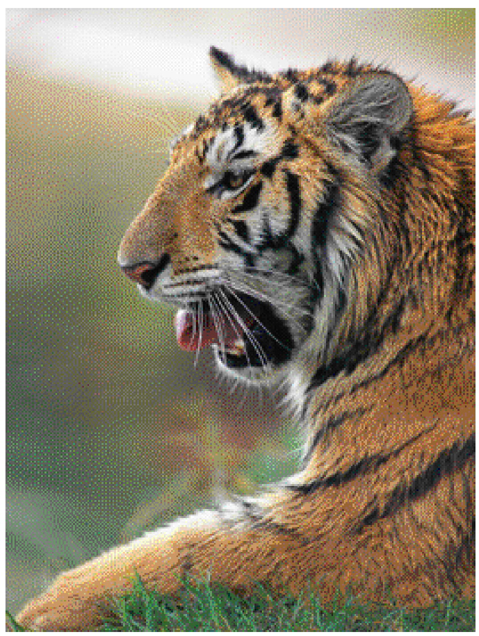
Mutualistic interactions between species can be of two types: ***symbiotic*** or ***nonsymbiotic***. In a ***symbiotic mutualism***, individuals interact physically and their relationship is biologically essential for survival. At least one member of the pair cannot live without close contact with the other. For example, the ***fungal-algal*** symbiosis that occurs in ***lichens***. The morphological structure of a lichen is a mass of fungal ***hyphae*** that forms around a small colony of algae cells. In this mutualism, the alga produces carbohydrates and other food by products through ***photosynthesis*** and ***metabolism***, while the fungus absorbs the required minerals and water to allow for these processes to occur.

More common in nature is the ***nonsymbiotic mutualism***. In this interaction, the mutualists live independent lives yet cannot survive without each other. The most obvious example of an interaction of this type is the relationship between flowering plants and their insect pollinators (**Figure 9f-1**).

{PRIVATE**}Figure 9f-1**: Bees and many species of flowering plants interact with each other in a mutualistic fashion. In this interaction, the flower becomes pollinated by the insect, while the bee receives food in the form of pollen and nectar.

Predation, Parasitism, and Pathogens

Pathogens, ***parasites***, and ***predators*** obtain food at the expense of their ***hosts*** and ***prey***. These processes are basic to the entire grazing food chain above the autotroph level. Predators tend to be larger than their prey and consume them from the outside (**Figure 9f-2**). A parasite or pathogen is smaller than its host and consumes it either from the inside or from the outside of the organism.



{PRIVATE**}Figure 9f-2**: The tiger (Panthera tigris) hunts at night preying on a variety of animals, including deer, wild hog, and wild cattle. Tigers are ambush predators that try to approaching their prey as closely as possible. They often attack their prey from behind, biting its neck or throat in the capture process.

It is easy to believe that the predator-prey interaction is somehow detrimental to the prey ***population***. This idea has led to extensive efforts to control predator populations in the name of wildlife conservation. However, functional relationships between predator-prey between species, within natural ecosystems, have coevolved over long periods time creating a dynamic balance between their interacting populations. Thus, the population sizes of predator and prey ***species*** are interregulated by delicate feedback mechanisms that control the densities of both species.

A classic example of the balance between predator and prey involves the prickly pear cactus, Opuntia spp. In the 19th century, prickly pear cactus was introduced into Australia from South America. Because no Australian predator species existed to control the population size of this cactus, it quickly expanded throughout millions of acres of grazing land. The presence of the prickly pear cactus excluded cattle and sheep from grazing vegetation and caused a substantial economic hardship to farmers. A method of control of the prickly pear cactus was initiated with the introduction of Cactoblastis cactorum, a cactus eating moth from Argentina, in 1925. By 1930, densities of the prickly pear cactus were significantly reduced.

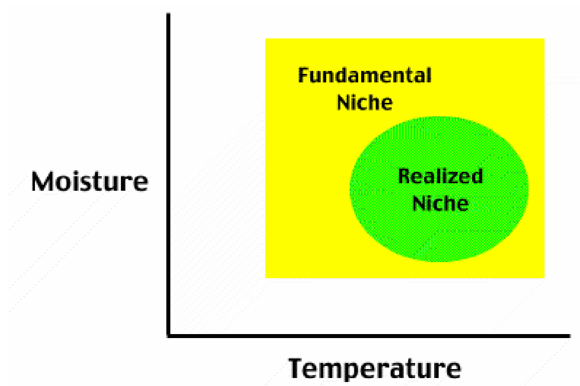
Sometimes predator species can drive their prey into localized ***extinction***. In complex communities, this does no particular harm to the predator if several other species exist as alternative prey.

(g) Concept of Ecological Niche

{PRIVATE**}**For a species to maintain its population, its individuals must survive and reproduce. Certain combinations of environmental conditions are necessary for individuals of each species to tolerate the physical environment, obtain energy and nutrients, and avoid predators. The total requirements of a species for all resources and physical conditions determine where it can live and how abundant it can be at any one place within its range. These requirements are termed abstractly the ***ecological niche***.

G.E. Hutchinson (1958) suggested that the niche could be modeled as an imaginary space with many dimensions, in which each dimension or axis represents the range of some environmental condition or ***resource*** that is required by the species. Thus, the niche of a plant might include the range of temperatures that it can tolerate, the intensity of light required for ***photosynthesis***, specific humidity regimes, and minimum quantities of essential soil nutrients for uptake.

A useful extension of the niche concept is the distinction between ***fundamental*** and ***realized niches*** (**Figure 9g-1**). The ***fundamental niche*** of a species includes the total range of environmental conditions that are suitable for existence without the influence of ***interspecific competition*** or ***predation*** from other species. The ***realized niche*** describes that part of the fundamental niche actually occupied by the species.



{PRIVATE**}Figure 9g-1:** The following diagram shows a hypothetical situation where a species distribution is controlled by just two environmental variables: temperature and moisture. The green and yellow areas describe the combinations of temperature and moisture that the species requires for survival and reproduction in its habitat. This resource space is known as the fundamental niche. The green area describes the actual combinations of these two variables that the species utilizes in its habitat. This subset of the fundamental niche is known as the realized niche.

(h) Species Diversity and Biodiversity

{PRIVATE}Biologists are not completely sure how many different ***species*** live on the Earth. Estimates of how many species exist on the Earth range from low of 2 million to high of about 100 million. To date, about 2.1 million species have been classified, primarily in the habitats of the middle latitudes. Most of the unclassified species on this planet are ***invertebrates***. This group of organisms includes insects, spiders, mollusks, sponges, flatworms, starfish, urchins, earthworms, and crustaceans. These species are often difficult to find and identify because of their small size and the fact that they live in habitats that are difficult to explore. In the tropical rain forest, the cataloging of species has been quite limited because of this later reason. Scientists estimate that this single biome may contain 50 to 90 % of the Earth's ***biodiversity***.

Many species have gone ***extinct*** over the Earth's geologic history. The primary reason for these extinctions is environmental change or biological competition. Since the beginning of the Industrial Revolution, a large number of biologically classified species have gone extinct due to the actions of humans. This includes 83 species of mammals, 113 species of birds, 23 species of amphibians and reptiles, 23 species of fish, about 100 species of invertebrates, and over 350 species of plants. Scientists can only estimate the number of unclassified species that have gone extinct. Using various methods of extrapolation, biologists estimate that in 1991 between 4000 to 50,000 unclassified species became extinct, mainly in the tropics, due to our activities. This rate of extinction is some 1,000 to 10,000 times greater than the natural rate of species extinction (2 - 10 species per year) prior to the appearance of human beings. The continued extinction of species on this planet by human activities is one of the greatest environmental problems facing humankind.

Several times during the Earth's history there have been periods of ***mass extinctions***, when many species became extinct in a relatively short time period (a few million years is a relatively short time when compared to the age of the Earth). Scientists are unsure of the causes of both background extinction and mass extinction. Possible explanations for mass extinctions include climate changes or catastrophes such as the Earth being hit by a meteor. Since the beginning of time, five or six mass extinctions have occurred that eliminated between 35 % and 96 % of all species on Earth (**Table 9h-1**). Further, it is believed that of all species that ever inhabited the Earth over 99 % of them are now extinct.

Table 9h-1: Major extinction events during the Phanerozoic.

{PRIVATE}**Date of the Extinction Event**	**Percent Species Lost**	**Species Affected**
65 Million Years Ago (Cretaceous)	85 %	Dinosaurs, plants (except ferns and seed bearing plants), marine vertebrates and invertebrates. Most mammals, birds, turtles, crocodiles, lizards, snakes, and amphibians were unaffected.
213 Million Years Ago (Triassic)	44 %	Marine vertebrates and invertebrates.
248 Million Years Ago (Permian)	75-95 %	Marine vertebrates and invertebrates.
380 Million Years Ago (Devonian)	70 %	Marine invertebrates.
450-440 Million Years Ago (Ordovician)	50 %	Marine invertebrates.

Assessment of the number of different organisms that live on this planet is plagued with difficulties. First and foremost, biologists lack a precise definition of what exactly defines a species. The concept of a species often refers to a population of physically similar individuals that can successfully mate between each other, but cannot produce fertile offspring with other organisms. However, many species are composed of a number of distinct populations that can interbreed even though they display physiological and anatomical differences. Scientists developed the notion of ***biodiversity*** to overcome some of the difficulties of species concept. To accomplish this task, biodiversity describes the diversity of life at the following three biological levels:

Genetic Level or **Genetic Diversity** - ***Genetic diversity*** refers to the total number of ***genetic*** characteristics expressed and recessed in all of the individuals that comprise a particular species

Species Level or **Species Diversity** - ***Species diversity*** is the number of different species of living things living in an area. As mentioned above, a species is a group of plants or animals that are similar and able to breed and produce viable offspring under natural conditions.

Ecosystem Level or **Ecosystem Diversity** - ***Ecosystem diversity*** is the variation of habitats, community types, and abiotic environments present in a given area. An ecosystem consists of all living and non-living things in a given area that interact with one another.

The biodiversity found on Earth today is the product of 3.5 billion years of evolution. In fact, the Earth supports more biodiversity today than in any other period in history. However, much of this biodiversity is now facing the threat of extinction because of the actions of humans.

(i) Plant Succession

{PRIVATE**}Introduction**

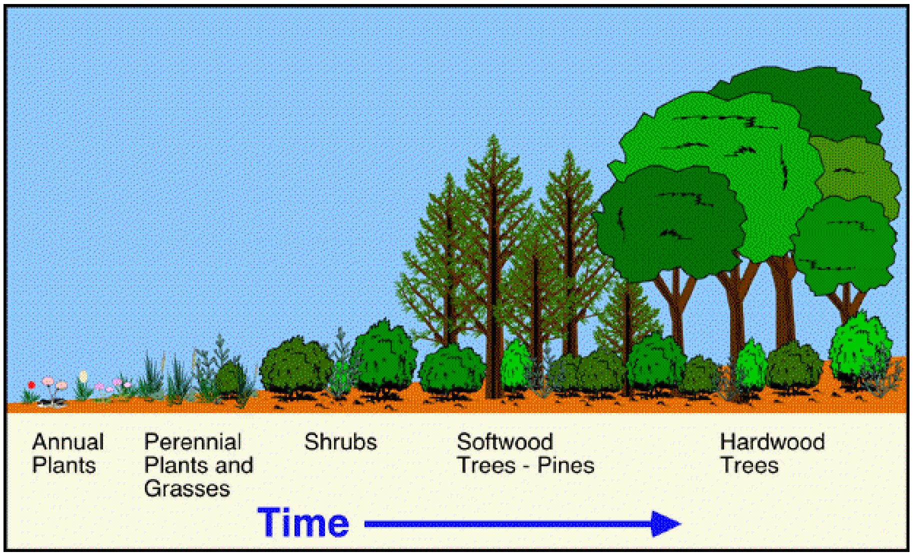
Succession is a directional non-seasonal cumulative change in the types of ***plant species*** that occupy a given area through time. It involves the processes of colonization, establishment, and extinction which act on the participating plant species. Most successions contain a number of stages that can be recognized by the collection of species that dominate at that point in the succession. Succession begin when an area is made partially or completely devoid of vegetation because of a ***disturbance***. Some common mechanisms of disturbance are fires, wind storms, volcanic eruptions, logging, climate change, severe flooding, disease, and pest infestation. Succession stops when species composition changes no longer occur with time, and this community is said to be a ***climax community***.

The concept of a climax community assumes that the plants colonizing and establishing themselves in a given region can achieve ***stable equilibrium***. The idea that succession ends in the development of a climax community has had a long history in the fields of biogeography and ecology. One of the earliest proponents of this idea was Frederic Clements who studied succession at the beginning of the 20th century. However, beginning in the 1920s scientists began refuting the notion of a climax state. By 1950, many scientists began viewing succession as a phenomenon that rarely attains equilibrium. The reason why equilibrium is not reached is related to the nature of ***disturbance***. Disturbance acts on communities at a variety of spatial and temporal scales. Further, the effect of disturbance is not always 100 percent. Many disturbances remove only a part of the previous plant community. As a result of these new ideas, plant communities are now generally seen as being composed of numerous patches of various size at different stages of successional development.

Abandoned Field to Oak Forest

One of the earliest studies of plant succession was done by Dwight Billings in the 1930s (see 1938, ***Ecological Monographs*** 8: 437-499). In this investigation, Billings examined the succession of plant species that occurred on abandoned agricultural fields in North Carolina. Billings studied a number of fields that had been deserted from just a few years to a maximum of about 150 years. From observations of the plant communities that existed in these sites, Billings was able to construct a detailed successional sequence.

The first stage of succession was characterized by the pioneering colonization of ***annual*** plant species on bare ground and nutrient poor soils (**Figure 9i-1**). These annual species had short lifespans (one growing season), rapid maturity, and produce numerous small easily dispersed seeds. The annuals were then quickly replaced in dominance in the next year by ***biennial*** plants and ***grasses***. After about 3 to 4 years, the biennial and grass species gave way to ***perennial herbs*** and ***shrubs***. These plants live for many years and have the ability to reproduce several times over their lifespans.



{PRIVATE}**Figure 9i-1:** Succession of plant species on abandoned fields in North Carolina. ***Pioneer species*** consist of a variety of annual plants. This successional stage is then followed by communities of perennials and grasses, shrubs, softwood trees and shrubs, and finally hardwood trees and shrubs. This succession takes about 120 years to go from the ***pioneer*** stage to the ***climax*** community.

After about 5 to 15 years, the sites were then colonized by a number of different softwood ***tree*** species including loblolly pine (Pinus taeda), shortleaf pine (Pinus echinata), Virginia pine (Pinus virginiana), and sweetgum. As the softwoods increased in their numbers and grew in height, they began forming a forest canopy. This canopy reduces the amount of light reaching the forest floor. The resulting shaded understory conditions caused the exclusion of many light loving perennial herb and shrub species. Low light conditions also inhibited the germination of pine seedlings. Perennial herb and shrub species that were adapted to low light conditions now began to take over the ground cover. The canopy also changed the microclimate of habitat near ground level. It was now more humid, has moderated temperatures, and less wind. These conditions, plus the development of a soil ***litter layer***, allowed for the germination of hardwood species, like oak (Quercus spp.) and various species of hickory (Carya spp). The seedlings of these tree species also tolerate low light levels.

By about 50 to 75 years after the initial colonization of the pioneer species, the hardwoods started to replace the softwood species in the developing forest. At this stage in the succession, the pines had maximum heights of about 25 meters, while the oaks and hickories were on average about 10 meters tall. Because of their shorter lifespans (50 years), many of the softwood species were beginning to die out and the gap that was created was then filled by a subdominant hardwood tree. Hardwood species, like oak and hickory, can live for more than 100 years. Sites more than 100 years old were found to be dominated by mature oak forests.

Organismic and Individualistic Views of Succession

In the first quarter of this century there was considerable debate about the nature of the ***community***. **F.E. Clements** (1916) conceived of the community as a sort of superorganism whose member ***species*** were tightly bound together both now and in their common evolutionary history. Thus, individuals, populations, and communities have a relationship to each other that resembles the associations found between cells, tissues, and organs.

Clements' theories on communities were also linked to ***succession***. His successional concept suggested that communities of organisms are subject to special laws in which the action of the whole is greater than the sum of the parts, and that this action results in a directional change in the species composition of the community to a ***climax*** state controlled primarily by climate.

The main processes acting to produce the various successional stages of species dominance, and finally climax, are competition and plant modification of the abiotic environment. Environmental modification, however, is detrimental to the plants doing the modifying. Modification changes the environment allowing the establishment of new colonists, and then results in the subsequent competitive exclusion of the former inhabitants by these colonists. This ***facilitative*** process stops when the climax community is reached.

Clements presented a **deterministic unidirectional view** of succession where the present pattern is governed by the past pattern. The philosophical structure of Clements' ***holistic*** approach is quite similar to advances in other sciences of that time. In geology and geomorphology theorists presented views that contained various stages of maturity, and compared landscape evolution metaphorically to a developing organism. The parallel in theoretical approach of these sciences with the work of Clements may be the result of attitudes prevalent in science as a whole at that time.

In 1926, **H. Gleason** (1926) devised a new theory to explain the nature of ***communities***. Gleason's **individualistic concept** saw the relationship between coexisting ***species*** as simply the result of similarities in their requirements and tolerances (and partly the result of chance). Taking this view, ***community boundaries*** need not be sharp, and associations of species would be much less predictable than one would expect from the superorganism concept.

Gleason argued that the ***holistic*** view point of Clements was inadequate in explaining the mechanism of succession. For example, Gleason suggested that Clementsian concepts could not properly explain the occurrence of such phenomena as ***retrogressive successions***. In reference to his view of succession, Gleason stated that ".... every species of plant is a law unto itself, the distribution of which in space depends upon its individual peculiarities of migration and environmental requirements". Thus, associations of plants, or communities, were not highly organized, but aggregations of independent plant species, each specialized to survive on habitats they were adapted for. ***Retrogressive successions*** were possible in Gleason's model if environmental variables deteriorated with time, changing the pattern of establishment, growth and reproduction of plants in a habitat. Clements' model, however, assumed long term climatic stability, and this assumption does not allow for short term retrogressive community change.

Clements and Gleason presented two **diametrically opposed** opinions on community organization and structure. Further investigation in this discussion will show that these views are still present in the hypotheses of later theorists, but in a somewhat modified form. Many of these modified hypotheses involve a synthesis of the early ideas of Clements and Gleason. This synthesis is the result of the addition of new ecological information or the re-analysis of old information on how ecosystems function over time. The synthetic evolution of successional hypotheses must be expected, as investigation finds new mechanisms responsible for temporal vegetation change in a relatively unexplored world. The early presence of simple diametrically opposed successional hypotheses in the early years is probably the result of the immature state of understanding of turn of the century ecology.

Central to Gleason's succession model is the notion of ***abiotic*** and ***biotic heterogeneity*** in space and time. This concept is a characteristic view of much of modern ecology. Recently, several scientists have examined the role of ***disturbance*** on community structure. These researchers suggest that disturbance is a common process in most communities that shapes the nature and structure of ***biotic interactions*** and processes. These ideas follow directly from Gleason's early observations of pattern and process in the plant community.

The **individualistic concept** of succession outlined by Gleason was ignored by the scientific community for some twenty to thirty years. Important papers and books citing this work did not appear until the late 1940s and early 1950s. It was the ideas of Clements that dominated ecological thought in one way or another up to this period.

Our current view on the nature of community structure is close to the **individualistic concept**. Results of many studies indicate that a given location, by virtue mainly of its physical characteristics, possesses a reasonable predictable association of species. However, a given species that occurs in one predictable association is also quite likely to occur with another group of species under different conditions elsewhere.

Types of Succession

Primary succession - is the establishment of plants on land that has not been previously vegetated - Mount Saint Helens. Begins with ***colonization*** and ***establishment*** of ***pioneer species***.

Secondary succession - is the invasion of a habitat by plants on land that was previously vegetated. Removal of past vegetation may be caused by natural or human ***disturbances*** such as fire, logging, cultivation, or ***hurricane***.

Allogenic succession - is caused by a change in environmental conditions which in turn influences the composition of the plant community. There have been a number of reports of an apparent allogenic transition between salt-marsh and woodland. The estuary of the River Fal in Cornwall, England, like many other estuaries, is subject to a quite rapid deposition of silt. This accretion can occur at a rate of 1 cm per year on the mud flats which are found 15 kilometers into the estuary. As a result, salt marsh has extended 800 meters seawards during the last century, while valley woodland has kept pace by invading the landward limit of the marshland.

Autogenic succession - is a succession where both the plant community and environment change, and this change is caused by the activities of the plants over time. Mt. St. Helens after the last volcanic eruption.

Progressive succession - is a succession where the community becomes complex and contains more species and biomass over time.

Retrogressive succession - is a succession where the community becomes simplistic and contains fewer species and less biomass over time. Some retrogressive successions are allogenic in nature. For example, the introduction of grazing animals result in degenerated rangeland.

Table 9i-1 describes some of the plant, community, and ecosystem attributes that change with succession.

{PRIVATE**}Table 9i-1:** Comparison of plant, community, and ecosystem characteristics between early and late stages of succession

{PRIVATE**}Attribute**	**Early Stages of Succession**	**Late Stages of Succession**
Plant Biomass	Small	Large
Plant Longevity	Short	Long
Seed Dispersal Characteristics of Dominant Plants	Well dispersed	Poorly dispersed
Plant Morphology and Physiology	Simple	Complex
Photosynthetic Efficiency of Dominant Plants at Low Light	Low	High
Rate of Soil Nutrient Resource Consumption by Plants	Fast	Slow
Plant Recovery Rate from Resource Limitation	Fast	Slow
Plant Leaf Canopy Structure	Multilayered	Monolayer
Site of Nutrient Storage	Litter and Soil	Living Biomass and Litter
Role of Decomposers in Cycling Nutrients to Plants	Minor	Great
Biogeochemical Cycling	Open and Rapid	Closed and Slow
Rate of Net Primary Productivity	High	Low
Community Site Characteristics	Extreme	Moderate (Mesic)
Importance of Macroenvironment on Plant Success	Great	Moderate
Ecosystem Stability	Low	High
Plant Species Diversity	Low	High
Life-History Type	r	K
Seed Longevity	Long	Short

Succession Mechanisms

An overview of the mechanisms of succession has been produced by Connell and Slatyer (1977, ***American Naturalist*** 111: 1119-1144). Connell and Slatyer propose three models, of which the first (***facilitation***) is the classical explanation most often invoked in the past, while the other two (***tolerance*** and ***inhibition***) may be equally important but have frequently been overlooked.

The essential feature of ***facilitation succession***, in contrast with either the tolerance or inhibition models, is that changes in the ***abiotic environment*** are imposed by the developing ***plant community***. Thus, the entry and growth of the later species depends on earlier species preparing the ground.

The ***tolerance model*** suggests that a predictable sequence is produced because different species have different strategies for exploiting ***resources***. Later species are able to tolerate lower resource levels due to ***competition*** and can grow to maturity in the presence of early species, eventually out competing them.

The ***inhibition model*** applies when all species resist invasions of competitors. Later species gradually accumulate by replacing early individuals when they die. An important distinction between models is the cause of death of the early colonists. In the case of facilitation and tolerance, they are killed in competition for resources, notably light and nutrients. In the case of the inhibition model, however, the early species are killed by very local ***disturbances*** caused by extreme physical conditions or the action of ***predators***.

(j) Introduction to the Ecosystem Concept

{PRIVATE}Introduction

In **topic *9d***, an ***ecosystem*** was defined as a dynamic entity composed of a biological ***community*** and its associated ***abiotic*** environment. Often the dynamic interactions that occur within an ecosystem are numerous and complex. Ecosystems are also always undergoing alterations to their biotic and abiotic components. Some of these alterations begin first with a change in the ***state*** of one component of the ecosystem which then cascades and sometimes amplifies into other components because of relationships.

In recent years, the impact of humans has caused a number of dramatic changes to a variety of ecosystems found on the Earth. Humans use and modify natural ecosystems through agriculture, forestry, recreation, urbanization, and industry. The most obvious impact of humans on ecosystems is the loss of ***biodiversity***. The number of ***extinctions*** caused by human domination of ecosystems has been steadily increasing since the start of the ***Industrial Revolution***. The frequency of species extinctions is correlated to the size of human population on the Earth which is directly related to resource consumption, land-use change, and environmental degradation. Other human impacts to ecosystems include species invasions to new habitats, changes to the abundance and dominance of species in communities, modification of ***biogeochemical cycles***, modification of hydrologic cycling, pollution, and climatic change.

Major Components of Ecosystems

Ecosystems are composed of a variety of abiotic and biotic components that function in an interrelated fashion. Some of the more important components are: **soil**, **atmosphere**, **radiation from the sun**, **water**, and **living organisms**.

Soils are much more complex than simple sediments. They contain a mixture of weathered rock fragments, highly altered soil mineral particles, ***organic matter***, and living organisms. Soils provide ***nutrients***, water, a home, and a structural growing medium for organisms. The vegetation found growing on top of a soil is closely linked to this component of an ecosystem through nutrient cycling.

The **atmosphere** provides organisms found within ecosystems with carbon dioxide for ***photosynthesis*** and oxygen for ***respiration***. The processes of ***evaporation***, ***transpiration***, and ***precipitation*** cycle water between the atmosphere and the Earth's surface.

Solar radiation is used in ecosystems to heat the atmosphere and to ***evaporate*** and ***transpire*** water into the atmosphere. Sunlight is also necessary for ***photosynthesis***. Photosynthesis provides the energy for plant growth and metabolism, and the organic food for other forms of life.

Most living tissue is composed of a very high percentage of **water**, up to and even exceeding 90 %. The ***protoplasm*** of a very few cells can survive if their water content drops below 10 %, and most are killed if it is less than 30-50 %. Water is the medium by which mineral nutrients enter and are translocated in plants. It is also necessary for the maintenance of leaf turgidity and is required for photosynthetic chemical reactions. Plants and animals receive their water from the Earth's surface and soil. The original source of this water is precipitation from the atmosphere.

Ecosystems are composed of a variety of **living organisms** that can be classified as ***producers***, ***consumers***, or ***decomposers***. ***Producers*** or ***autotrophs***, are organisms that can manufacture the organic compounds they use as sources of energy and ***nutrients***. Most producers are green plants that can manufacture their food through the process of ***photosynthesis***. ***Consumers*** or ***heterotrophs*** get their energy and nutrients by feeding directly or indirectly on producers. We can distinguish two main types of consumers. ***Herbivores*** are consumers that eat plants for their energy and nutrients. Organisms that feed on herbivores are called ***carnivores***. Carnivores can also consume other carnivores. Plants and animals supply organic matter to the soil system through shed tissues and death. Consumer organisms that feed on this organic matter, or ***detritus***, are known as ***detritivores*** or ***decomposers***. The organic matter that is consumed by the detritivores is eventually converted back into ***inorganic*** nutrients in the soil. These nutrients can then be used by plants for the production of organic compounds.

The following **graphical** model describes the major ecosystem components and their interrelationships (**Figure 9j-1**).

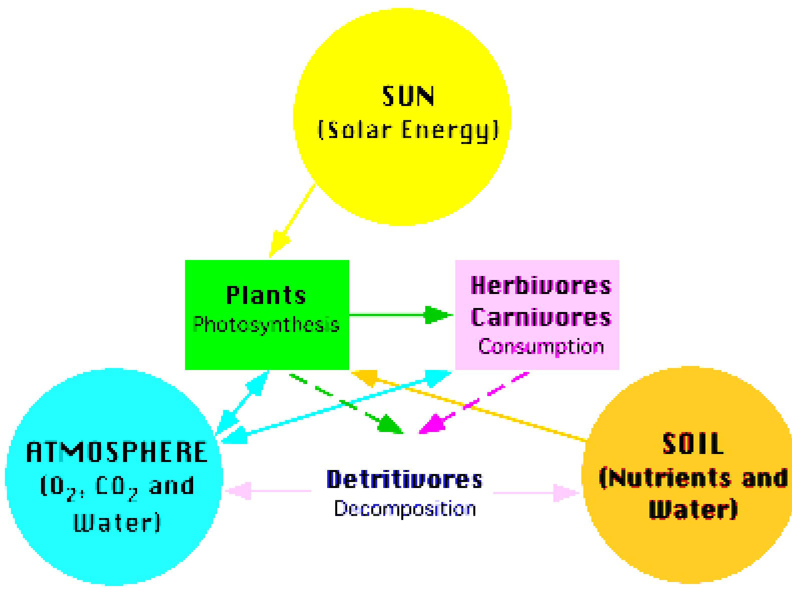


Figure 9j-1: Relationships within an ecosystem.

Energy and Matter Flow in Ecosystems

Many of the most important relationships between living organisms and the environment are controlled ultimately by the amount of available incoming energy received at the Earth's surface from the sun. It is this energy which helps to drive biotic systems. The sun's energy allows plants to convert ***inorganic*** chemicals into ***organic*** compounds.

Only a very small proportion of the sunlight received at the Earth's surface is transformed into biochemical form. Several studies have been carried out to determine this amount. A study of an Illinois cornfield reported that 1.6 % of the available solar radiation was photosythetically utilized by the corn. Other data suggests that even the most efficient ecosystems seldom incorporate more than 3 % of the available solar insolation. Most ecosystems ***fix*** less than 1 % of the sunlight available for ***photosynthesis***.

Living organisms can use energy in basically two forms: ***radiant*** or ***fixed***. **Radiant energy** exists in the form of ***electromagnetic energy***, such as light. **Fixed energy** is the ***potential chemical energy*** found in organic substances. This energy can be released through ***respiration***. Organisms that can take energy from inorganic sources and fix it into energy rich organic molecules are called ***autotrophs***. If this energy comes from light then these organisms are called ***photosynthetic autotrophs***. In most ecosystems plants are the dominant photosynthetic autotroph.

Organisms that require fixed energy found in organic molecules for their survival are called ***heterotrophs***. Heterotrophs who obtain their energy from living organisms are called ***consumers***. Consumers can be of two basic types: Consumer and decomposers. Consumers that consume plants are know as ***herbivores***. ***Carnivores*** are consumers who eat herbivores or other carnivores. ***Decomposers*** or ***detritivores*** are heterotrophs that obtain their energy either from dead organisms or from organic compounds dispersed in the environment.

Once fixed by plants, organic energy can move within the ecosystem through the consumption of living or dead organic matter. Upon decomposition the chemicals that were once organized into organic compounds are returned to their inorganic form and can be taken up by plants once again. Organic energy can also move from one ecosystem to another by a variety of processes. These processes include: ***animal migration***, animal harvesting, plant harvesting, plant ***dispersal*** of ***seeds***, ***leaching***, and ***erosion***. The following **diagram** models the various inputs and outputs of energy and matter in a typical ecosystem (**Figure 9j-2**).

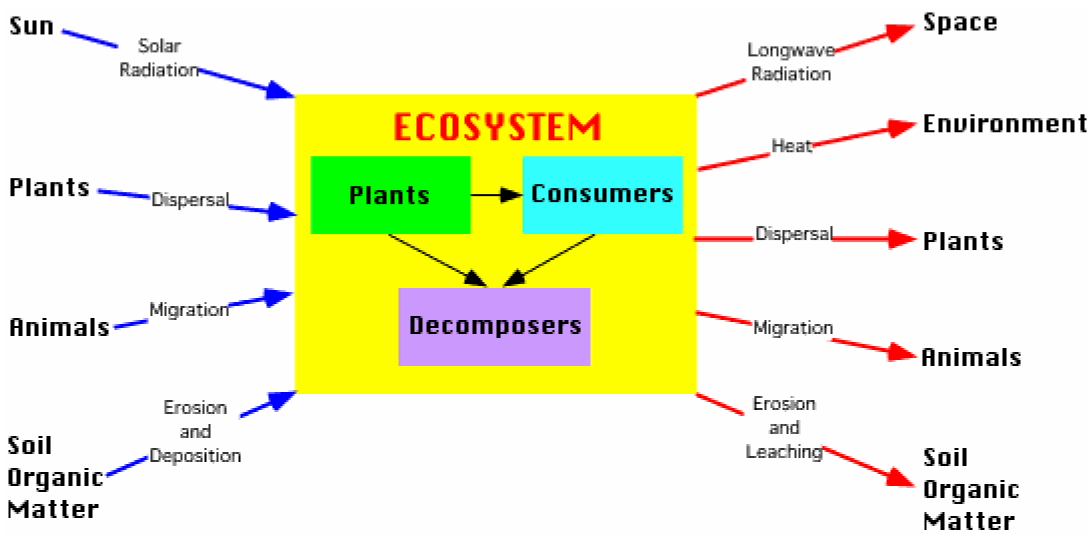
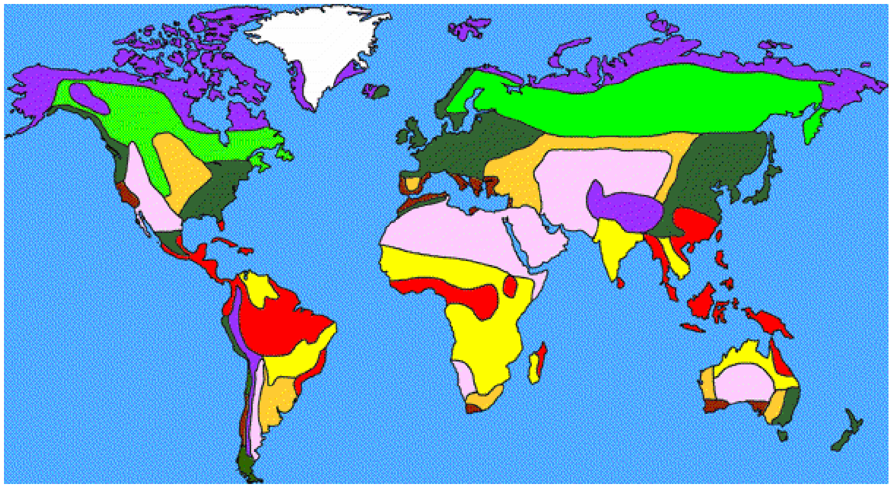


Figure 9j-2: Inputs and outputs of energy and matter in a typical ecosystem.

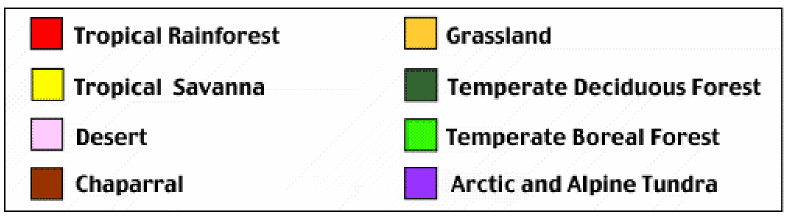
(k) Characteristics of the Earth's Terrestrial Biomes

{PRIVATE}Introduction

Many places on Earth share similar climatic conditions despite being found in geographically different areas. As a result of ***natural selection***, comparable ***ecosystems*** have developed in these separated areas. Scientists call these major ecosystem types ***biomes***. The geographical distribution (and ***productivity***) of the various biomes is controlled primarily by the climatic variables ***precipitation*** and ***temperature***. The map in **Figure 9k-1** describes the geographical locations of the eight major biomes of the world. Because of its scale, this map ignores the many community variations that are present within each biome category.



{PRIVATE**}Figure 9k-1:** Distribution of the Earth's eight major terrestrial biomes. Legend is below. (**Adapted** from: H.J. de Blij and P.O. Miller. 1996. Physical Geography of the Global Environment. John Wiley, New York. Pp. 290.)



Most of the classified biomes are identified by the dominant plants found in their communities. For example, ***grasslands*** are dominated by a variety of ***annual*** and ***perennial*** species of ***grass***, while ***deserts*** are occupied by plant species that require very little water for survival or by plants that have specific adaptations to conserve or acquire water.

The ***diversity*** of ***animal*** life and subdominant plant forms characteristic of each biome is generally controlled by abiotic environmental conditions and the productivity of the dominant vegetation. In general, ***species diversity*** becomes higher with increases in ***net primary productivity***, moisture availability, and temperature.

Adaptation and ***niche specialization*** are nicely demonstrated in the biome concept. Organisms that fill similar niches in geographically separated but similar ecosystems usually are different species that have undergone similar ***adaptation*** independently, in response to similar environmental pressures. The vegetation of California, Chile, South Africa, South Australia, Southern Italy and Greece display similar morphological and physiological characteristics because of convergent evolution. In these areas, the vegetation consists of drought-resistant, hard-leaved, low growing woody shrubs and trees like eucalyptus, olive, juniper, and mimosa.

Arctic and Alpine Tundra

Tundra means marshy plain. The geographical distribution of the tundra biome is largely poleward of 60 degrees North latitude. The tundra biome is characterized by an absence of trees, the presence of dwarf plants, and a ground surface that is wet, spongy, and hummocky. Soils of this biome are usually permanently frozen (***permafrost***) starting at a depth of a few centimeters to meter or more. The permafrost line is a physical barrier to plant root growth.

Within this biome, temperature, precipitation, and evaporation all tend to be at a minimum. Most tundra locations, have summer months with an average temperature below 10 degrees Celsius. Precipitation in the wettest month is normally no higher that 25 millimeters. However, despite the low levels of precipitation the ground surface of the tundra biome is often waterlogged because of low rates of ***evapotranspiration***.

The ***species diversity*** of tundra vegetation is relatively small. Plant communities are usually composed of a few species of dwarf ***shrubs***, a few ***grass*** species, ***sedges***, and ***mosses***. Perhaps the most characteristic arctic tundra plants are ***lichens*** like Reindeer Moss (Cladonia spp.). The principal ***herbivores*** in this biome include caribou, musk ox, arctic hare, voles, and lemmings. Most of the bird species of the tundra have the ability to migrate and live in warmer locations during the cold winter months. The herbivore species support a small number of ***carnivore*** species like the arctic fox, snow owl, polar bear, and wolves. ***Reptiles*** and ***amphibians*** are few or completely absent because of the extremely cold temperatures.

Alpine tundra is quite similar to some arctic tundra but differs in the absence of permafrost and in the presence of better drainage.

Boreal Coniferous Forest

This moist-cool, transcontinental ***coniferous*** forest, or ***taiga*** lies largely between the 45th and 57th North latitudes. The climate of this biome is cool to cold with more precipitation than the tundra, occurring mainly in the summer because of mid-latitude cyclones. The predominant vegetation of boreal biome are needle-leaf evergreen variety ***tree*** species. Some common species include: White Spruce (Picea glauca) and Balsam Fir (Abies balsamea) east of the Rockies; Red Pine (Pinus resinosa), White Pine (Pinus strobus), and Hemlock (Tsuga canadensis) in the Great Lakes Region. In British Columbia, dominant boreal trees include: Black Spruce (Picea mariana), White Spruce (Picea glauca), Lodgepole Pine (Pinus contorta), Ponderosa Pine (Pinus pondersoa), Douglas Fir (Pseudotsuga menziesii), and Alpine Fir (Abies lasiocarpa). The understory is relatively limited as a result of the low light penetration even during the spring and fall months. Understory plants in the deciduous biome take advantage of the leafless condition of trees during these seasons concentrating there growth during this time period. Common understory species include orchids, ***shrubs*** like rose, blueberry, and cranberry (**Figure 9k-2**).

Figure 9k-2: The understory of boreal forest habitats is usually poorly developed.

Mammals common to the boreal forest include moose, bear, deer, wolverine, marten, lynx, wolf, snowshoe hare, vole, chipmunks, shrews, and bats (**Figure 9k-3**). ***Reptiles*** are rare, once again, because of cold temperatures .

Figure 9k-3: Bears are common in the boreal forest ecosystem.

Boreal forest ***soils*** are characterized by a deep ***litter layer*** and slow ***decomposition***. Soils of this biome are also ***acidic*** and mineral deficient because of the large movement of water vertically though the profile and subsequent ***leaching***.

Temperate Deciduous Forest

As its name indicates, this biome is characterized by a moderate climate and ***deciduous*** trees. It once occupied much of the eastern half of the United States, central Europe, Korea, and China. This biome has been very extensively affected by human activity, and much of it has been converted into agricultural fields or urban developments. Dominant plants include trees like Maple (Acer spp.), Beech (Fagus spp.), Oak (Quercus spp.), Hickory (Carya spp.), Basswood (Tilia spp.), Cottonwood (Populus spp.), Elm (Ulmus spp.), and Willow (Salix spp.). The understory of shrubs and herbs in a mature deciduous forest is typically well developed and richly diversified. Many different types of ***herbivores*** and ***carnivores***, and some ***reptiles*** and ***amphibians*** exist here.

Brown forest soils characterize temperate deciduous forest ecosystems. The surface ***litter layer*** in these soils is thin due to rapid ***decomposition***.

Grassland

In central North America are the ***grasslands***, the tall grass prairie toward the east and the short grass prairie westward. In Europe and Asia some grasslands are called Steppes. In South America, grasslands are known as Pampas. Prior to modern man, the tall grass prairie was dominated by species of Bluestem (Andropogon spp.) (**Figure 9k-4**). This particular species dominated much of the tall grass prairie forming dense covers 1.5 to 2.0 meters tall. In the western end of the prairie, where precipitation is less, Buffalo Grass (Buchloe dactyloides) and other grasses only a few inches above the soil surface are common in this habitat. Flowering ***herbs***, including many kinds of ***composites*** and ***legumes***, are common but much less important than grass species. Trees are limited to low lying areas and the narrow zone immediately adjacent to streams.

{PRIVATE**}Figure 9k-4:** Natural grassland ecosystems are dominated by various species of grass. (**Source**: NASA - Oak Ridge National Laboratory (***ORNL***) Distributed Active Archive Center (***DAAC***) *Net Primary Productivity (NPP) Database*).

Figure 9k-5: Much of the prairies is now cultivated to grow grain crops.

In the tall grass prairie organic rich and black ***chernozemic*** soils are common. Chernozems are among the richest in ***nutrients*** and consequently the most fertile in the world. In drier parts of prairies, soils can be influenced by ***salinization***. As a result of their fertility, most grassland ecosystems have been modified by humans to grow grain and other dryland crops (**Figure 9k-5**).

Grassland ***mammals*** are dominated by smaller burrowing herbivores (prairie dogs, jack rabbits, ground squirrels, and gophers) and larger running herbivores such as bison, pronghorn antelope, and elk. ***Carnivores*** include badger, coyote, ferret, wolf, and cougar. The population size of many of these species has been drastically reduced because of habitat destruction (**Figure 9k-5**). Some of these species are on the edge of ***extinction***.

Desert

In its most typical form, the desert consists of ***shrub*** covered land where the plants are spatially quite dispersed. The major desert biomes of the Earth are geographically found at between 25 to 35 degrees North and South latitude, in the interiors of continents. Climatically, deserts are influenced by descending air currents which limit the formation of precipitation. Many desert areas have less than 25 millimeters of precipitation annually. Dominant plants include drought resistant shrubs like the Creosote Bush (Larrea divaricata) and Sagebrush (Artemisia tridentata), water storing ***succulents*** like cactus (**Figure 9k-6**), and many species are short lived ***annuals*** that complete their life cycles during infrequent and short rainy periods. Desert habitats can be devoid of vegetation if precipitation is in very short supply (**Figure 9k-7**).

Figure 9k-6: Cactus are a common type of drought resistant plant found in deserts.

Figure 9k-7: Desert habitat devoid of vegetation.

Most desert ***mammals*** tend to be nocturnal to avoid the high temperatures. Desert habitats have a rich lizard and snake fauna because high temperatures promote the success of cold blooded life forms (**Figure 9k-8**).

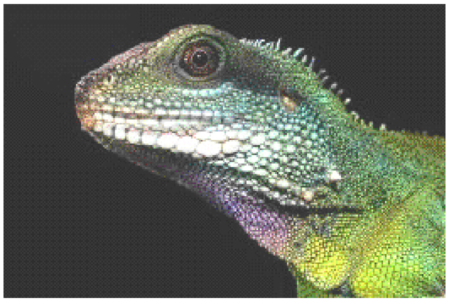


Figure 9k-8: Lizards are quite common in desert habitats.

Because productivity is low, the ***litter layer*** is comparably limited and organic content of surface soil layers is very low. Also, evaporation tends to concentrate salts at the soil surface.

Chaparral

Chaparral has a very specific spatial distribution. It is found in a narrow zone between 32 and 40 degrees latitude North and South on the west coasts of the continents. This area has a dry climate because of the dominance of the ***subtropical high pressure zone*** during the fall, summer, and spring months. ***Precipitation*** falls mainly in the winter months because of the seasonal movement of the ***polar front*** and its associated ***mid-latitude cyclone*** storms. Annual averages range from about 300 to 750 millimeters and most of this rain falls in a period between 2 to 4 months long. As a result of the climate, the vegetation that inhabits this biome exhibits a number of adaptations to withstand drought and fire. Trees and shrubs living in this zone tend to be small with hard evergreen leaves. Plants in the chaparral do not drop their leaves during the dry season because of the expense of replacement. The dry climate slows the rate leave decomposition in the soil. As a result, the plants growing in this biome do not have nutrients available for uptake to produce new leaves when the wet season begins. Instead, the plants of the chaparral develop leaves that can withstand arid conditions.

Representative species of the chaparral include cork oak (Quercus suber), olive (Olea europaea), eucalyptus, arbutus (Arbutus unedo), acacia, maritime pine (Pinus pinaster), shrub oak (Quercus dumosa), and live oak (Quercus virginiana). Many of the plant species have thorns to protect them from herbivore damage.

This biome is sometimes also called Mediterranean Scrubland or sclerophyll forest.

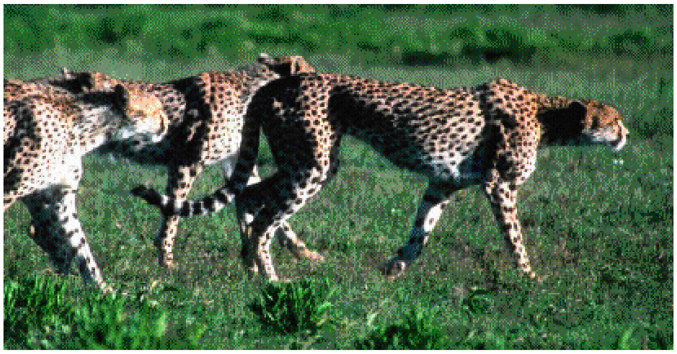
Tropical Savanna

Tropical savannas are ***grasslands*** with scattered drought-resistant ***trees*** that generally do not exceed 10 meters in height (**Figure 9k-9**). Tree and shrub species in the savanna usually shed their leaves during the dry season. This adaptation reduces water loss from the plants. New leaves appear several weeks before the start of the rain season. Scientists believe that savanna plant species may have developed this strategy to take advantage of the season variance of the start of the rains. Climatically, these biomes are characterized by distinct wet and dry seasons. Temperatures are hot all year long.

{PRIVATE}**Figure 9k-9:** Savanna vegetation is typical composed of a mixture of grass and trees. (**Source**: NASA - Oak Ridge National Laboratory (***ORNL***) Distributed Active Archive Center (***DAAC***) *Net Primary Productivity (NPP) Database*).

The savanna biome constitutes extensive areas in eastern Africa, South America, and Australia. Savannas also support the richest ***diversity*** of grazing ***mammals*** in the world (**Figure 9k-10**). The grazing animals provide food for a great variety of predators (**Figure 9k-11**).

Figure 9k-10: Zebras are a common grazer of the African savanna.



{PRIVATE}**Figure 9k-11:** Savannas are also home to a number of predator species who prey on grazing animals.

The soils are more ***nutrient*** rich than tropical forest soils. Some soils become extremely dry because of evaporation and form ***laterite*** layers.

Tropical Rainforest

Tropical rainforests occur in a broad zone outside the ***equator***. Annual rainfall, which exceeds 2000 to 2250 millimeters, is generally evenly distributed throughout the year. Temperature and humidity are relatively high through the year. Flora is highly diverse: a square kilometer may contain as many as 100 different tree species as compared to 3 or 4 in the temperate zone. The various trees of the tropical rain forests are closely spaced together and form a thick continuous canopy some 25 to 35 meters tall (**Figure 9k-12**). Every so often this canopy is interrupted by the presence of very tall trees (up to 40 meters) that have wide buttressed bases for support. ***Epiphytic*** orchids and ***bromeliads***, as well as vines (***lianas***), are very characteristic of the tropical rainforest biome. Some other common plants include ***ferns*** and palms. Most plants are evergreen with large, dark green, leathery leaves.

{PRIVATE**}Figure 9k-12:** Tropical trees often have buttressed bases to help support their heavy above-ground biomass. (**Source**: NASA - Oak Ridge National Laboratory (***ORNL***) Distributed Active Archive Center (***DAAC***) *Net Primary Productivity (NPP) Database*).

The tropical rainforest is also home to a great variety of animals (**Figure 9k-13**). Some scientists believe that 30 to 50 % of all of the Earth's animal species may be found in this biome.

Figure 9k-13: The tropical rain forest is home to many different species of amphibians.

Decomposition is rapid in the tropicals because high temperatures and an abundance of moisture. Because of the frequent and heavy rains, tropical soils are subject to extreme ***chemical weathering*** and ***leaching***. These environmental conditions also make tropical soils ***acidic*** and ***nutrient*** poor.

(l) Primary Productivity of Plants

{PRIVATE**}Introduction**

The bodies of living organisms within a unit area make up a standing crop of ***biomass***. More specifically, biomass can be defined as the mass of organisms per unit area and is usually expressed in units of energy (e.g., joules m^{-2}) or dry organic matter (e.g., tons ha^{-1} or grams m^{-2}). Most of the biomass in a ***community*** is composed of plants, which are the ***primary producers*** of biomass because of their ability to ***fix*** carbon through ***photosynthesis***. This chemical reaction can be described by the following simple formula:

$6CO_2 + 6H_2O$ + light energy = $C_6H_{12}O_6 + 6O_2$

The product of photosynthesis is a ***carbohydrate***, such as the sugar ***glucose***, and oxygen which is released into the atmosphere (**Figure 9l-1**). All of the sugar produced in the photosynthetic cells of plants and other organisms is derived from the initial chemical combining of carbon dioxide and water with sunlight (**Figure 9l-1**). This chemical reaction is catalyzed by ***chlorophyll*** acting together with other ***pigment***, ***lipid***, ***sugar***, ***protein***, and ***nucleic acid*** molecules. Sugars created in photosynthesis can be later converted by the plant to starch for storage, or it can be combined with other sugar molecules to form specialized carbohydrates, such as ***cellulose***. Sugars can also be combined with other ***nutrients*** such as nitrogen, phosphorus, and sulfur, to build complex molecules such as proteins and nucleic acids.

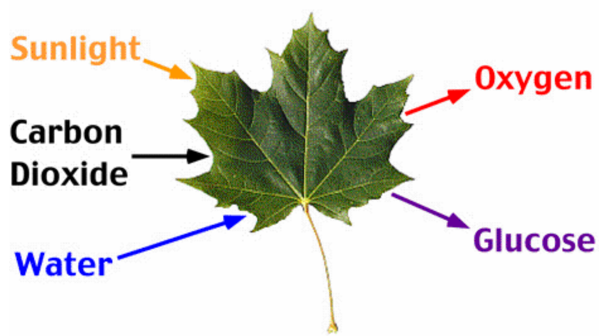


Figure 9l-1: Inputs and outputs of the photosynthetic process.

The **primary productivity** of a ***community*** is the amount of biomass produced through photosynthesis per unit area and time by plants, the primary producers. Primary productivity is usually expressed in units of energy (e.g., joules m^{-2} day^{-1}) or in units of dry organic matter (e.g., kg m^{-2} $year^{-1}$). Globally, primary production amounts to 243 billion metric tons of dry plant biomass per year. The total energy ***fixed*** by plants in a community through photosynthesis is referred to as ***gross primary productivity*** (**GPP**). Because all the energy fixed by the plant is converted into sugar, it is theoretically possible to determine a plant's energy uptake by measuring the amount of sugar produced. A proportion of the energy of gross primary productivity is used by plants in a process called ***respiration***. Respiration provides a plant with the energy needed for various plant physiological and morphological activities. The general equation for respiration is:

$C_6H_{12}O_6 + 6O_2 = 6CO_2 + 6H_2O$ + released energy

Subtracting respiration from gross primary production gives us ***net primary productivity*** (**NPP**), which represents the rate of production of biomass that is available for consumption (***herbivory***) by ***heterotrophic*** organisms (bacteria, fungi, and animals).

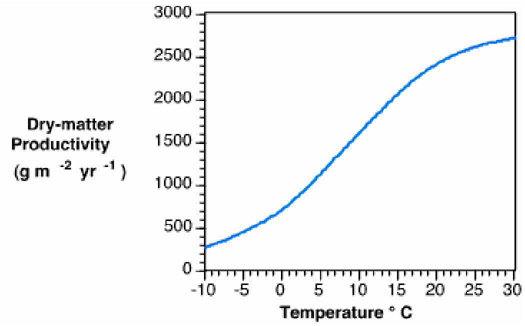
Globally, patterns of primary productivity vary both spatially and temporally. The least productive ecosystems are those limited by ***heat energy*** and water like the deserts and the polar tundra. The most productive ecosystems are systems with high temperatures, plenty of water and lots of available soil nitrogen. **Table 9l-1** describes the approximate average net primary productivity for a variety of ecosystem types.

Table 9l-1: Average annual Net Primary Productivity of the Earth's major biomes.

{PRIVATE}**Ecosystem Type**	**Net Primary Productivity** (kilocalories/meter^{-2}/year)
Tropical Rain Forest	9000
Estuary	9000
Swamps and Marshes	9000
Savanna	3000
Deciduous Temperate Forest	6000
Boreal Forest	3500
Temperate Grassland	2000
Polar Tundra	600
Desert	< 200

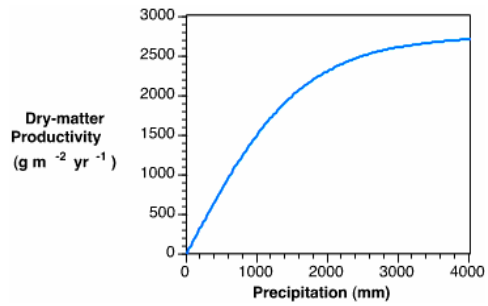
Factors Limiting Primary Productivity

Although all biological activity in plants is ultimately dependent on received solar radiation, it is obvious that solar radiation alone does not determine primary productivity. All plants require sunlight, carbon dioxide, and water for photosynthesis. Photosynthesis is also dependent on temperature and soil ***nutrients***. Temperature (heat) controls the rate of plant ***metabolism*** which in turn determines the amount of photosynthesis that can take place. Most biological metabolic activity takes place within the range 0 to 50 degrees Celsius. There is little activity above or below this range. The optimal temperatures for productivity coincide with 15 to 25 degrees Celsius optimal range of photosynthesis. The graph in the **Figure 9l-2** illustrates the relationship between the net primary productivity of forests with annual air temperature.



{PRIVATE}**Figure 9l-2:** Relationship between forest net primary productivity and annual temperature. (**Adapted** from: H. Lieth. 1973. Primary production: terrestial ecosystems. Human Ecology 1: 303-332.)

The general relationship between net primary productivity and ***precipitation*** for forests of the world is shown in the following **Figure 9l-3**.



{PRIVATE}**Figure 9l-3:** Relationship between forest net primary productivity and annual precipitation. (**Adapted** from: H. Lieth. 1973. Primary production: terrestial ecosystems. Human Ecology 1: 303-332.)

Water is a principal requirement for photosynthesis and the main chemical component of most plant cells. In dry regions, there is a linear increase in NPP with increased water availability. In the more humid forest climates of the world, plant productivity begins to level off at higher levels of precipitation.

Efficiency of Solar Radiation Use

Depending on location, between 0 and 5 ***joules*** of solar energy is received on each square meter of the Earth's surface every minute. However, only a small proportion of this energy is converted by photosynthesis into plant biomass. Of the sunlight being received at the Earth's surface, only about 44 % of the incident ***shortwave radiation*** occurs at wavelengths (known as **photosynthetically active radiation**, PAR) useful for photosynthesis. Yet, even the most efficient plant species (most of which are crop plants) can only incorporate 3-10 % of photosynthetically active radiation into the production of biomass. Of the Earth's various biomes, tropical rainforest and conifer forest are the most efficient converting between 1 to 3 % of the useable solar radiation into biomass. Deciduous forests achieve photosynthetic efficiencies between 0.5 to 1.0 %. The desert biome has the lowest efficiencey of solar radiation use. The plants in this biome convert only 0.01 to 0.2 % of photosynthetically active radiation to biomass.

(m) Production by Consumers and the Grazing Food Chain

{PRIVATE**}Production by Consumers**

Biological communities also include organisms that consume biomass for their nutrition. Such organisms include ***herbivores***, ***carnivores***, and ***detritivores***. These organisms obtain their energy through ***respiration***, a process that releases energy from organic molecules like ***glucose***. The equation for respiration is described below:

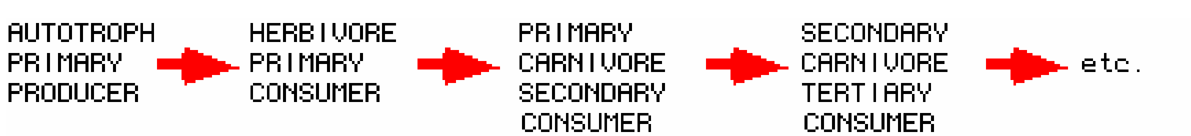
$C_6H_{12}O_6 + 6O_2 = 6CO_2 + 6H_2O$ + released energy

The amount of energy actually used by these animal populations is significantly less than the amount consumed. In all animals, digestion is an imperfect process and only a portion of the energy ingested is actually ***assimilated*** and then used for body maintenance, growth and reproduction. The remaining portion leaves the organism's body undigested.

We sometimes call the energy assimilated by ***consumer*** organisms the ***gross secondary productivity***. Gross secondary productivity can be determined directly, unlike ***gross primary productivity***, by measuring the amount consumed minus the material defecated.

The Grazing Food Chain

The ***grazing food chain*** is a model that describes the general flow of energy in communities. For most ecosystems the model begins with the photosynthetic fixation of light, carbon dioxide, and water by plant ***autotrophs*** (***primary producers***) who produce sugars and other organic molecules. Once produced, these compounds can be used to create the various types of plant tissues. ***Primary consumers*** or ***herbivores*** form the second link in the grazing food chain. They gain their energy by consuming primary producers. ***Secondary consumers*** or ***primary carnivores***, the third link in the chain, gain their energy by consuming herbivores. ***Tertiary consumers*** or ***secondary carnivores*** are animals that receive their organic energy by consuming primary carnivores. The illustation below models this process:



The various levels in the grazing food chain are linked to each other like links in a chain. The levels are often called ***trophic levels***, and they suggest a particular order for the passage of energy through the food chain. Like many very simple models, the idea of a food chain only provides a simple abstraction of the nature of energy flow through communities. The ultimate disposition of the energy assimilated by consumers is by four routes:

respiration, biomass accumulation, decay of organic matter by bacteria and other ***decomposer*** organisms, and consumption by consumers.

The actual amount of energy incorporated in the tissues of consumers at each tropic level is not determined by the gross amounts consumed. Rather it is the amount of organic energy converted into actual biomass. Consumers lose significant amounts of consumed energy due to ***assimilation*** inefficiencies, morphological and physiological maintenance, reproduction, and the process of finding and capturing food. The energy to perform the latter three processes is supplied by ***respiration***.

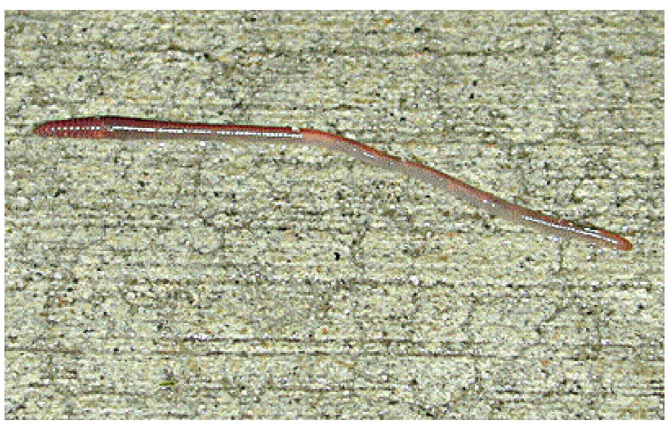
Thus, the number of ***trophic levels*** that can be maintained in any ecosystem is finite. The limit is reached when consumers can no longer consume enough energy to balance the energy lost in the assimilation process, morphological and physiological maintanence, growth, reproduction, and in finding the food. Normally, ecosystems have about **four** or **five** trophic levels.

Finally, some ecosystems, like rivers and lakes, are characterized by large imports of biomass in the form of dead organic matter. This situation can cause consumer production to be higher than the ***autotrophs*** found in these systems.

(n) Organic Decomposition and the Detritus Food Chain

{PRIVATE**}**The ***detritus food chain*** differs from the grazing food chain in several significant ways. First, the organisms making it up are in general physically smaller. Second, the functional roles of the different organisms do not fall as neatly into categories like the grazing food chain's trophic levels. Finally, ***detritivores*** live in environments (soil, sea bed, etc.) that are rich in scattered food particles. As a result, ***decomposers*** are less mobile than ***herbivores*** or ***carnivores***.

The organisms of the detritus food chain include members of many different species of animals and plants, such as algae, bacteria, slime molds, fungi, protozoa, insects, mites, crustaceans, centipedes, mollusks, worms, sea cucumbers, and even some vertebrates (**Figure 9n-1**). These organisms consume organic wastes, shed tissues, and the dead bodies of both plants and animals.



{PRIVATE**}Figure 9n-1:** Earthworms are one of the most important soil decomposers. These organisms consume vast amounts of organic matter and mineral soil. As the organic matter passes through their digestive system, it is subjected to digestive enzymes and the grinding action of mineral soil particles. The amount of material consumed per day is often equal to their body weight.

Decomposers tend to always be active, processing large amounts of ***organic matter*** and releasing a great deal of energy mostly as heat from metabolic activities (e.g., compost heap). The end result of decomposition is the conversion of organic matter back into its original ***inorganic nutrient*** form. In mature forest and grassland soils, the ***decomposition*** process establishes an equilibrium over time where ***litterfall*** additions equal the amount of organic matter decomposed.

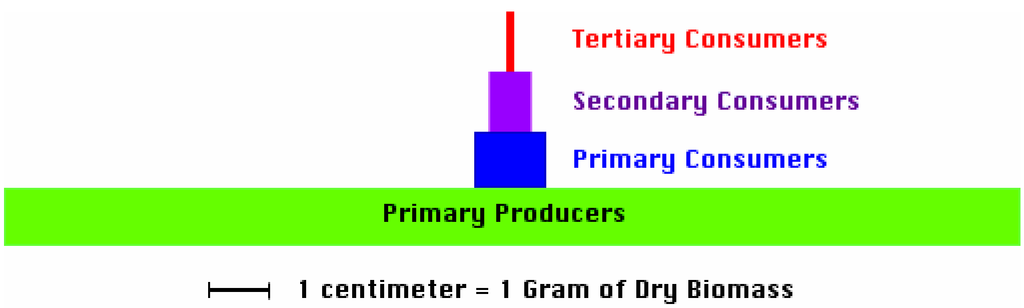
The rate of ***decomposition*** in a detritus food chain is controlled by many factors. In most terrestrial ecosystems, temperature and soil oxygen and moisture content tend to be the primary variables controlling the activities of decomposers. In some ecosystems, oxygen may not be readily available. In these cases ***aerobic respiration*** cannot take place, and the breakdown of organic compounds and energy extraction must then proceed by ***anaerobic*** means like ***fermentation***. Organisms involved in fermentation include bacteria and yeast. **Anaerobic** decomposition releases much less energy from organic matter than does **aerobic** respiration. The end products of anaerobic decomposition include molecules such as carbon dioxide, water, and humus. They also include small-molecule alcohols, organic acids, ptomaines, amines, and other products, as well as gaseous substances such as methane.

Because respiration is so much more efficient at releasing the energy contained in organic molecules, the activity of the detritus food chain is much higher in an aerobic environment, and the breakdown of materials more complete. Organic matter breakdown is substantially slower and less complete in anaerobic environments. It also results in the accumulation of undegraded organic matter in the form of ***peat***, organic soils, and highly organic sediments.

In conclusion, there is no waste in a properly functioning natural ecosystem. Everything once living or alive will be consumed by decomposers at some time and returned to inorganic form.

(o) Trophic Pyramids and Food Webs

{PRIVATE}So far we have described food chains as morphological systems of energy flow. The energy flow within food chains can also be described in more quantitative terms. Several different quantitative models are commonly seen in the academic literature. One of these models, called a ***pyramid of biomass***, quantifies all of the living ***biomass*** found in each of the ***trophic levels***. Biomass can be defined as the weight of living matter (usually measured in dry weight per unit area). **Figure 9o-1** describes the pyramid of biomass for an aquatic community living in a shallow experimental pond.



{PRIVATE}**Figure 9o-1:** Pyramid of biomass for a pond. (**Source:** Data from Whittaker, R.H. 1961. Experiments with radiophosphorus tracer in aquarium microcosms. **Ecological Monographs** 31:157-188).

In most ecosystems, the amount of biomass found in each trophic level decreases progressively as one moves from the beginning to the end of the grazing food chain. As described in previous sections, ***primary producers*** or **plants** are the original source of fixed organic energy in ecosystems. The energy that plants fix supports the life found in all other trophic levels. However, only the ***primary consumers*** or ***herbivores*** directly feed on the primary producers.

The amount of organic energy incorporated into the biomass of the primary consumers is significantly smaller than the amount found in the primary producer level. Theoretically, herbivores could consume all available plant life within an ecosystem. However, in reality this rarely occurs because plants have developed, through evolution, a number of mechanisms that protect most of their tissues from consumption. As a result, herbivores can only consume a portion of the available plant biomass. All that is consumed does not become herbivore biomass. Significant losses of biomass occur because of digestive inefficiencies and ***respiration***. ***Assimilation*** efficiencies for most terrestrial herbivores range from 20 to 60 percent. Some of the assimilated biomass is lost through the process of **respiration.**

The next level in the pyramid of biomass is the ***secondary consumers*** or ***primary carnivores***. These organisms harvest a portion of the herbivore biomass for their nutrition. Once again, not all herbivores are eaten because of defensive mechanisms. Most herbivores possess some evolutionary adaptation that generally protects them from carnivore attack. These adaptations include the ability to fly and run, body armor, quills and protective spines, and camouflage. In general, carnivores have higher assimilation efficiencies than herbivores. Their assimilation efficiencies range from 50 to 90 percent. Only a portion of the assimilated organic energy becomes carnivore biomass because of the metabolic energy needs of body maintenance, growth, reproduction, and locomotion.

Many food chains have no more than four or five trophic levels. In the example above, the studied ecosystem had four. This final level is composed of the ***tertiary consumers*** who feed on the ***secondary consumers***. The amount of biomass found in this trophic level is very small relative to the other levels. This is to be expected because of the processes, as described above, that cause continuously less energy to be available to successive consumers.

Trophic pyramids have also been constructed to show the transfer of energy in caloric terms and the number of organisms found in each trophic level.

A model describing the organisms found in a food chain is called a ***food web*** (**Figure 9o-2**). Food webs describe the complex patterns of energy flow in an ecosystem by modeling who consumes who. The **illustration** below describes a portion of the food web for a typical tidal marsh ecosystem located on the southern coast of British Columbia.

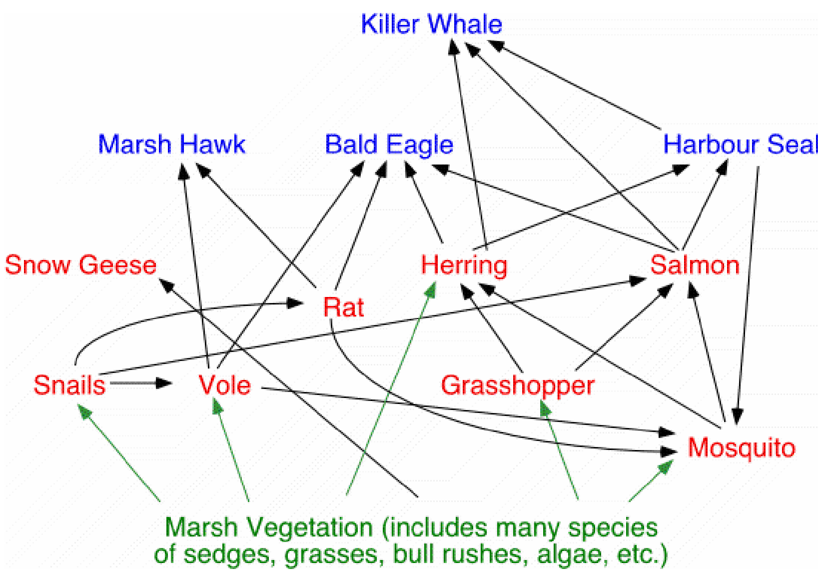


Figure 9o-2: Typical tidal marsh food web.

(p) Biogeochemical Cycling: Inputs and Outputs of Nutrients to Ecosystems

{PRIVATE**}**The patterns of cycling ***nutrients*** in the ***biosphere*** involves not only ***metabolism*** by living organisms, but also a series of strictly ***abiotic*** chemical reactions. Understanding the cycle of a single ***element*** requires the knowledge of a process that depends jointly on the biology of all organisms that utilize the element, its geological availability, and its organic and inorganic chemistry. Thus, understanding the cycling of biologically important elements is truly interdisciplinary in nature. We generally call this process ***biogeochemical cycling***.

All biogeochemical cycles for all ecologically significant ***elements*** have both an ***organic*** and an ***inorganic*** component. Both are extremely important! How efficiently the nutrients move through the organic component back to the inorganic reserviors determines how much is available to organisms over the short term. The major reservoirs for all metabolically important elements are found either in the ***atmosphere***, ***lithosphere*** (mainly ***rock***,

soil and other ***weathered sediments***) or ***hydrosphere***. Flow in the inorganic phase generally tends to be slower than in the organic phase.

Nutrients Essential for Life

Living organisms require the availability of about 20 to 30 chemical elements for the various of ***metabolic*** processes that take place in their bodies. Some products of this metabolism require relatively few nutrients for their production. For example, ***carbohydrates*** are photosynthesized from just water and carbon dioxide. Some organic substances, like ***amino acids*** and ***proteins***, are more complex in their chemical make up and therefore require a number of different ***nutrients***.

The types of nutrient needed by life is often categorized into two groups. Elements required in relatively large amounts are generally referred to as ***macronutrients***. Macronutrients that constitute more than 1 % each of dry weight include **carbon**, **oxygen**, **hydrogen**, **nitrogen**, and **phosphorus**. Macronutrients that constitute 0.2 to 1 % of dry organic weight include **sulfur**, **chlorine**, **potassium**, **sodium**, **calcium**, **magnesium**, **iron**, and **copper**.

Nutrients needed in trace amounts are generally called ***micronutrients***. These elements often constitute less than 0.2 % of dry organic matter. Some common micronutrients required by living organisms include **aluminum**, **boron**, **bromine**, **chromium**, **cobalt**, **fluorine**, **gallium**, **iodine**, **manganese**, **molybdenum**, **selenium**, **silicon**, **strontium**, **tin**, **titanium**, **vanadium**, and **zinc**.

Nutrient Inputs to Ecosystems

Important ***nutrients*** for life generally enter ***ecosystems*** by way of four processes:

(1). Weathering

Rock weathering is one of the most important long-term sources for nutrients. However, this process adds nutrients to ecosystems in relatively small quantities over long periods of time. Important nutrients released by weathering include:

Calcium, magnesium, potassium, sodium, silicon, iron, aluminum, and phosphorus.

All of the micro nutrients.

Carbon, oxygen and nitrogen are not inputted into ***ecosystems*** by weathering. The main source for these important elements is the atmosphere and the decomposition of ***organic matter***.

(2). Atmospheric Input

Large quantities of nutrients are added to ecosystems from the ***atmosphere***. This addition is done either through ***precipitation*** or by a number of **biological processes**.

Carbon - absorbed by way of ***photosynthesis*.**

Nitrogen - produced by ***lightning*** and ***precipitation***.

Sulfur, **chloride**, **calcium**, and **sodium** - deposited by way of ***precipitation***.

The quantity of nitrogen added to ***ecosystems*** by lightning and rain annually ranges from 1 to 20 kilograms per hectare depending on geographical location. A value of 5 to 8 kilograms per hectare is typical for temperate ecosystems like deciduous forest or grasslands.

(3). Biological Nitrogen Fixation

Biological ***nitrogen fixation*** is the biochemical process where elemental nitrogen is combined into organic forms by metabolic processes. It is carried out by a limited number of organisms including several species of ***bacteria*** (not all are associated with legumes), a few **actinomycetes** (***fungi***), and **blue-green algae** (***cyanobacteria***).

The amount of nitrogen fixed by biological process has been estimated to be quite large. Recent calculations suggest that total annual global fixation of nitrogen is about 175 million metric tons.

Symbiotic Fixation with Legumes

The symbiosis of ***legumes*** (13,000 species) and bacteria of the genus Rhizobium provides the major biological source of fixed nitrogen in agricultural soils. Rhizobium bacteria invade the root hairs and the cortical cells, ultimately inducing the formation of nodules that serve as a home for the organisms. The host plant supplies the bacteria with carbohydrates for energy, and the bacteria reciprocate by supplying the plant with fixed nitrogen compounds, an association that is ***mutually*** beneficial.

The nitrogen fixed by Rhizobium goes in three directions. First, it is used directly by the host plant. Second, as the nitrogen passes from plant roots and nodules into the soil, some of it is ***mineralized*** and becomes available almost immediately as ***ammonium*** or ***nitrate*** compounds. The final pathway for the fixed nitrogen is that it can be incorporated into the bodies of the general purpose decay organisms and finally into the ***soil organic matter***.

Symbiotic Fixation with Non-Legumes

About 160 species of non-legumes are known to develop nodules to house ***bacteria*** for ***nitrogen fixation***. Some of the commonly known plant species that have this ability belong to the alder and birch family.

Symbiotic Nitrogen Fixation Without Nodules

Recent studies have called attention to several significant nonlegume symbiotic associations that do not produce plant nodules. Among the most important are those involving **blue-green algae** (***cyanobacteria***).

(4). Immigration

The immigration of motile animals into an ecosystem can sometimes add significant additions of nutrients to an ecosystem that are locked up in the biomass of the organisms. These nutrients are released when the organism dies or sheds its tissues.

Nutrient Outputs to Ecosystems

Important ***nutrients*** required for life leave ***ecosystems*** by way of four processes:

(1). Erosion

Soil erosion is probably the most import means of nutrient loss to ecosystems. Erosion is very active in agricultural and forestry systems, where cultivation, grazing, and clearcutting leaves the soil bare and unprotected. When unprotected, the surface of the soil is easily transported by wind and moving water. The top most layers of a soil, which have an abundance of nutrient rich organic matter, are the major storehouse for soil nutrients like phosphorus, potassium, and nitrogen.

(2). Leaching

Another important process of nutrient loss is ***leaching***. Leaching occurs when water flowing vertically through the ***soil*** transports nutrients in solution downward in the soil profile. Many of these nutrients can be completely lost from the soil profile if carried into ***groundwater*** and then horizontally transported into ***rivers***, ***lakes***, or ***oceans***. Leaching losses are, generally, highest in disturbed ecosystems. In undisturbed ***ecosystems***, efficient nutrient cycling limits the amount of nutrients available for this process.

(3). Gaseous Losses

High losses of nutrients can also occur when specific environmental conditions promote the export of nutrients in a gaseous form. When the ***soil*** is wet and ***anaerobic***, many compounds are chemically ***reduced*** to a gas from solid forms in the soil. This is especially true of soil nitrogen. Scientific studies in Netherlands have shown that about 80 % of the nitrogen fertilizer applied to the soil for crop consumption may be lost through the process of ***denitrification***.

(4). Emigration and Harvesting

Just as material may be introduced to ecosystems by migration, so too may it be lost. The emigration of animals, and the removal of vegetation by humans are both processes by which outputs can occur from an ecosystem.

(q) Soil Organic Matter Decomposition and Nutrient Cycling

{PRIVATE}In terrestrial ***ecosystems*** most of the nutrient cycling occurs in the topmost horizons in ***soil***. The main **inputs** to the soil come from ***weathering***, rainfall, fertilizers, ***atmospheric*** sources and organisms. Under natural conditions, inputs from plants are the most important, including not only nutrients released by ***organic decomposition***, but also substances washed in from the plant leaves (***foliar leaching***). Losses or **outputs** are by ***leaching***, ***erosion***, gaseous loss (like ***denitrification***) and plant uptake. Within the soil, nutrients are stored on the soil particles, dead ***organic matter***, or in chemical compounds.

Organic matter decomposition is the main process that recycles nutrients back into the soil. Decomposition of organic matter begins with large soil organisms like **earthworms**, **arthropods** (ants, beetles, and termites), and **gastropods** (slugs and snails). These organisms breakdown the organic matter into smaller pieces which can be decomposed by smaller organisms like ***fungi*** and heterotrophic ***bacteria*** (**Figure 9q-1**).

{PRIVATE}**Figure 9q-1**: Fungi play an important role in the decomposition process converting organic matter back into basic inorganic chemicals.

Decomposition of organic matter may take several months to several years to complete. In tropical regions, the whole process is quite quick because moist conditions and high temperatures enhance biological activity.

(r) The Carbon Cycle

{PRIVATE}All life is based on the element **carbon**. Carbon is the major chemical constituent of most organic matter, from ***fossil fuels*** to the complex molecules (***DNA*** and ***RNA***) that control genetic reproduction in organisms. Yet by weight, carbon is not one of the most abundant elements within the Earth's crust. In fact, the lithosphere is only 0.032 % carbon by weight. In comparison, oxygen and silicon respectively make up 45.2 % and 29.4 % of the Earth's surface rocks.

Carbon is stored on our planet in the following major ***sinks*** (**Figure 9r-1** and **Table 9r-1**): (1) as ***organic molecules*** in living and dead organisms found in the ***biosphere***; (2) as the gas ***carbon dioxide*** in the ***atmosphere***; (3) as ***organic matter*** in ***soils***; (4) in the ***lithosphere*** as ***fossil fuels*** and ***sedimentary rock*** deposits such as ***limestone***, ***dolomite*** and ***chalk***; and (5) in the oceans as dissolved atmospheric ***carbon dioxide*** and as ***calcium carbonate*** shells in marine organisms.

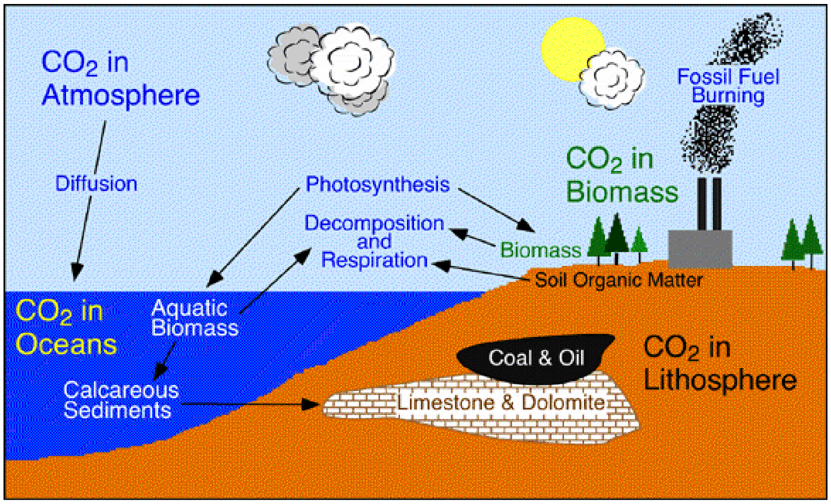


Figure 9r-1: Carbon cycle.

Table 9r-1: Estimated major stores of carbon on the Earth.

{PRIVATE}Sink	**Amount in Billions of Metric Tons**
Atmosphere	578 (as of 1700) - 766 (as of 1999)
Soil Organic Matter	1500 to 1600
Ocean	38,000 to 40,000
Marine Sediments and Sedimentary Rocks	66,000,000 to 100,000,000
Terrestrial Plants	540 to 610
Fossil Fuel Deposits	4000

Ecosystems gain most of their ***carbon dioxide*** from the ***atmosphere***. A number of ***autotrophic*** organisms have specialized mechanisms that allow for absorption of this gas into their cells. With the addition of water and energy from ***solar radiation***, these organisms use ***photosynthesis*** to chemically convert the carbon dioxide to carbon-based sugar molecules. These molecules can then be chemically modified by these organisms through the metabolic addition of other elements to produce more complex compounds like ***proteins***, ***cellulose***, and ***amino acids***. Some of the ***organic matter*** produced in plants is passed down to ***heterotrophic*** animals through consumption.

Carbon dioxide enters the waters of the ocean by simple ***diffusion***. Once dissolved in seawater, the carbon dioxide can remain as is or can be converted into carbonate (CO_3^{-2}) or bicarbonate (HCO_3^-). Certain forms of sea life biologically fix bicarbonate with calcium (Ca^{+2}) to produce ***calcium carbonate*** ($CaCO_3$). This substance is used to produce shells and other body parts by organisms such as coral, clams, oysters, some protozoa, and some algae. When these organisms die, their shells and body parts sink to the ocean floor where they accumulate as carbonate-rich deposits. After long periods of time, these deposits are physically and chemically altered into ***sedimentary rocks***. Ocean deposits are by far the biggest sink of carbon on the planet (**Table 9r-1**).

Carbon is released from ecosystems as ***carbon dioxide*** gas by the process of ***respiration***. Respiration takes place in both plants and animals and involves the breakdown of carbon-based organic molecules into carbon dioxide gas and some other compound by products. The ***detritus food chain*** contains a number of organisms whose primary ecological role is the ***decomposition*** of organic matter into its abiotic components.

Over the several billion years of geologic history, the quantity of carbon dioxide found in the atmosphere has been steadily decreasing. Researchers theorized that this change is in response to an increase in the sun's output over the same time period. Higher levels of carbon dioxide helped regulate the Earth's temperature to levels slightly higher than what is perceived today. These moderate temperatures allowed for the flourishing of plant life despite the lower output of ***solar radiation***. An enhanced ***greenhouse effect***, due to the greater concentration of carbon dioxide gas in the atmosphere, supplemented the production of ***heat energy*** through higher levels of longwave counter-radiation. As the sun grew more intense, several biological mechanisms gradually locked some of the atmospheric carbon dioxide into ***fossil fuels*** and ***sedimentary rock***. In summary, this regulating process has kept the Earth's global average temperature essentially constant over time. Some scientists suggest that this phenomena is proof for the ***Gaia hypothesis***.

Carbon is stored in the lithosphere in both ***inorganic*** and ***organic*** forms. Inorganic deposits of carbon in the ***lithosphere*** include ***fossil fuels*** like ***coal***, ***oil***, and ***natural gas***, oil ***shale***, and ***carbonate*** based sedimentary deposits like ***limestone***. Organic forms of carbon in the lithosphere include ***litter***, ***organic matter***, and ***humic*** substances found in soils. Some carbon dioxide is released from the interior of the lithosphere by ***volcanoes***. Carbon dioxide released by volcanoes enters the lower lithosphere when carbon-rich sediments and sedimentary rocks are ***subducted*** and partially melted beneath ***tectonic*** boundary zones.

Since the ***Industrial Revolution***, humans have greatly increased the quantity of carbon dioxide found in the Earth's atmosphere and oceans. Atmospheric levels have increased by over 30 %, from about 275 parts per million (ppm) in the early 1700s to just over 365 PPM today. Scientists estimate that future atmospheric levels of carbon dioxide could reach an amount between 450 to 600 PPM by the year 2100. The major sources of this gas due to human activities include fossil fuel combustion and the modification of natural plant cover found in grassland, woodland, and forested ecosystems. Emissions from fossil fuel combustion account for about 65 % of the additional carbon dioxide currently found in the Earth's atmosphere. The other 35 % is derived from deforestation and the conversion of natural ecosystems into agricultural systems. Researchers have shown that natural ecosystems can store between 20 to 100 times more carbon dioxide than agricultural land-use types.

(s) The Nitrogen Cycle

{PRIVATE}The ***nitrogen cycle*** represents one of the most important nutrient cycles found in terrestrial ecosystems (**Figure 9s-1**). Nitrogen is used by living organisms to produce a number of complex ***organic*** molecules like ***amino acids***, ***proteins***, and ***nucleic acids***. The largest store of nitrogen is found in the atmosphere where it exists as a gas (mainly N_2). The atmospheric store is about one million times larger than the total nitrogen contained in living organisms. Other major stores of nitrogen include organic matter in soil and the oceans. Despite its abundance in the atmosphere, nitrogen is often the most limiting nutrient for plant growth. This problem occurs because most plants can only take up nitrogen in two solid forms: ***ammonium*** ion (NH_4^+) and the ion ***nitrate*** (NO_3^-). Most plants obtain the nitrogen they need as ***inorganic*** nitrate from the ***soil solution***. Ammonium is used less by plants for uptake because in large concentrations it is extremely toxic. Animals receive the required nitrogen they need for ***metabolism***, growth, and reproduction by the consumption of living or dead organic matter containing molecules composed partially of nitrogen.

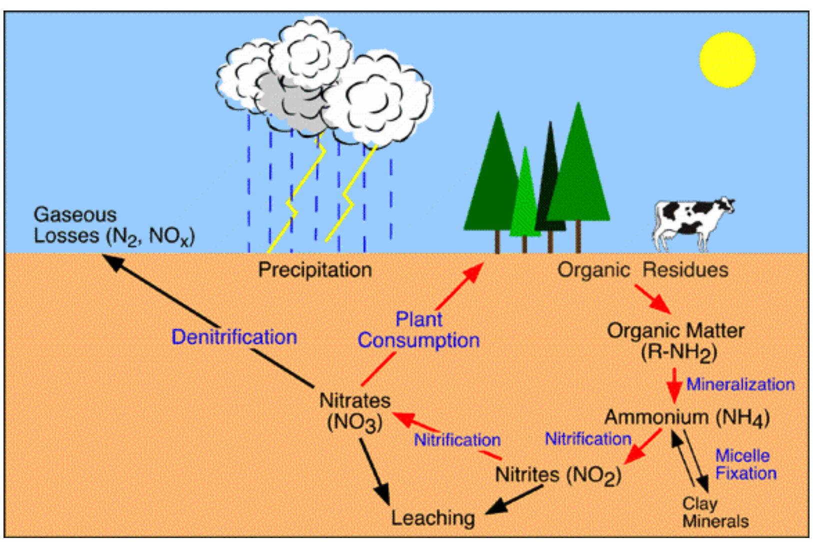


Figure 9s-1: Nitrogen cycle.

In most ecosystems nitrogen is primarily stored in living and dead ***organic matter***. This organic nitrogen is converted into inorganic forms when it re-enters the ***biogeochemical cycle*** via ***decomposition***. ***Decomposers***, found in the upper soil layer, chemically modify the nitrogen found in ***organic matter*** from ***ammonia*** (NH_3) to ***ammonium*** salts (NH_4^+). This process is known as ***mineralization*** and it is carried out by a variety of ***bacteria***, ***actinomycetes***, and ***fungi***.

Nitrogen in the form of ***ammonium*** can be absorbed onto the surfaces of clay particles in the soil. The ion of ammonium has a positive molecular charge is normally held by ***soil colloids***. This process is sometimes called **micelle fixation** (see **Figure 9s-1**). Ammonium is released from the colloids by way of ***cation exchange***. When released, most of the ammonium is often chemically altered by a specific type of ***autotrophic bacteria*** (bacteria that belong to the genus **Nitrosomonas**) into ***nitrite*** (NO_2^-). Further modification by another type of bacteria (belonging to the genus **Nitrobacter**) converts the ***nitrite*** to ***nitrate*** (NO_3^-). Both of these processes involve chemical ***oxidation*** and are known as ***nitrification***. However, nitrate is very soluble and it is easily lost from the soil system by ***leaching***. Some of this leached nitrate flows through the ***hydrologic system*** until it reaches the oceans where it can be returned to the atmosphere by ***denitrification***. Denitrification is also common in ***anaerobic*** soils and is carried out by ***heterotrophic bacteria***. The process of denitrification involves the ***reduction*** of nitrate into nitrogen (N_2) or nitrous oxide (N_2O) gas. Both of these gases then ***diffuse*** into the atmosphere. This process is important to the bacteria because it supplies them with oxygen for ***respiration***.

Almost all of the nitrogen found in any terrestrial ecosystem originally came from the atmosphere. Small proportions enter the soil in rainfall or through the effects of lightning. The majority, however, is biochemically ***fixed*** within the soil by specialized micro-organisms like ***bacteria***, ***actinomycetes***, and ***cyanobacteria***. Members of the bean family (legumes) and some other kinds of plants form mutualistic symbiotic relationships with nitrogen fixing bacterial. In exchange for some nitrogen, the bacteria receive from the plants carbohydrates and special structures (nodules) in roots where they can exist in a moist environment. Scientist estimate that biological fixation globally adds approximately 140 million metric tons of nitogen to ecosystems every year.

The activities of humans have severely altered the nitrogen cycle. Some of the major processes involved in this alteration include:

The application of nitrogen fertilizers to crops has caused increased rates of denitrification and leaching of nitrate into ***groundwater***. The additional nitrogen entering the groundwater system eventually flows into streams, rivers, lakes, and estuaries. In these systems, the added nitrogen can lead to ***eutrophication***.

Increased deposition of nitrogen from atmospheric sources because of fossil fuel combustion and forest burning. Both of these processes release a variety of solid forms of nitrogen through combustion.

Livestock ranching. Livestock release a large amounts of ammonia into the environment from their wastes. This nitrogen enters the soil system and then the hydrologic system through leaching, groundwater flow, and ***runoff***.

Sewage waste and septic tank ***leaching***.

10) Introduction to Geology

(a) The Rock Cycle

{PRIVATE}The ***rock cycle*** is a general model that describes how various geological processes create, modify, and influence rocks (**Figure 10a-1**). This model suggests that the origin of all rocks can be ultimately traced back to the solidification of molten ***magma***. Magma consists of a partially melted mixture of elements and compounds commonly found in rocks. Magma exists just beneath the solid crust of the Earth in an interior zone known as the ***mantle***.

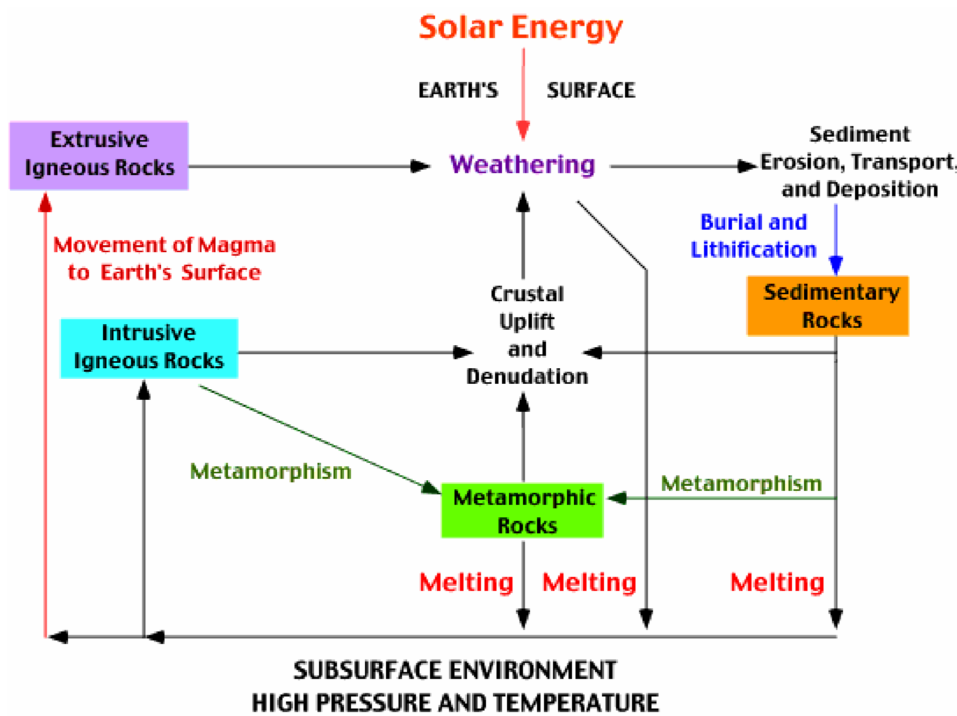


Figure 10a-1: The rock cycle.

Igneous rocks form from the cooling and crystallization of magma as it migrates closer to the Earth's surface. If the crystallization process occurs at the Earth's surface, the rocks created are called ***extrusive igneous rocks***. ***Intrusive igneous rocks*** are rocks that form within the Earth's solid ***lithosphere***. Intrusive igneous rocks can be brought to the surface of the Earth by ***denudation*** and by a variety of ***tectonic*** processes.

All rock types can be physically and chemically decomposed by a variety of surface processes collectively known as ***weathering***. The debris that is created by weathering is often ***transported*** through the landscape by ***erosional*** processes via streams, glaciers, wind, and gravity. When this debris is ***deposited*** as a permanent ***sediment***, the processes of burial, compression, and chemical alteration can modify these materials over long periods of time to produce ***sedimentary rocks***.

A number of geologic processes, like ***tectonic folding*** and ***faulting***, can exert heat and pressure on both igneous and sedimentary rocks causing them to be altered physically or chemically. Rocks modified in this way are termed ***metamorphic rocks***.

All of the rock types described above can be returned to the Earth's interior by ***tectonic*** forces at areas known as ***subduction zones***. Once in the Earth's interior, extreme pressures and temperatures melt the rock back into magma to begin the rock cycle again.

(b) Geologic Time

{PRIVATE}Geologists and geomorphologists describe the Earth's geologic history through a temporal system known as the ***geologic time scale*** (**Table 10b-1**). On this scale, time is measured using the following four units of time: ***eons***, ***eras***, ***periods*** and ***epochs***. All of these temporal subdivisions are established on the occurrence of some important geologic event. For example, **Hadean eon** represents the time on Earth when life did not exist. During the **Archean eon** life started and was dominated by one-celled ***prokaryotic*** life forms. ***Eukaryotic*** one-celled organisms became dominant in the **Proterozoic eon**. Multicellular organisms ruled the planet during the **eon** known as the ***Phanerozoic***.

Table 10b-1 describes some of the important geologic events that have occurred since the Earth's formation some 4.6 billion years ago.

Table 10b-1: Geologic time scale.

{PRIVATE}**Eon**	**Era**	**Period**	**Epoch**	**Major Geologic Milestones**
		Quaternary (0-1.6 million yrs BP)	***Holocene*** (Present-10,000 yrs BP)	Modern humans develop. Pleistocene Ice Age Interglacial.
			Pleistocene (10,000 -1,600,000 yrs BP)	Pleistocene Ice Age. Extinction of many species of large mammals and birds.
	Cenozoic	***Tertiary***	Pliocene (1.6-5.3 million yrs BP)	Development of hominid bipedalism. Cascade Mountains began forming. Climate cooling.
			Miocene (5.3-24 million yrs BP)	Chimpanzee and hominid lines evolve. Extensive glaciation in Southern Hemisphere. Climate cooling.
			Oligocene (24-37 million yrs BP)	Browsing mammals and many types of modern plants evolve. Creation of the Alps and Himalaya mountain chains. Volcanoes form in Rocky Mountains.
			Eocene (37-58 million yrs BP)	Primitive monkeys evolve and Himalayas began forming. Australian plate separates from Antarctica. Indian plate collides with Asia.
			Paleocene (58-65 million yrs BP)	Rats, mice, and squirrels evolve. Shallow continental seas become less widespread.
	Mesozoic	***Cretaceous*** (65-144 million yrs BP)		First flowering plants, greatest dinosaur diversity, Cretaceous Mass Extinction (65 m BP), and Andes Mountains form. Africa and South America begin to separate. Climate cooling because of mountain building. Shallow seas have extensive distribution.
		Jurassic (144-208 million yrs BP)		First birds and mammals appear. Nevadian Mountains form. Large areas of the continents covered by shallow seas. Climate generally warm and stable with little seasonal or latitudinal variation. Shallow seas expanding.
		Triassic (208-245 million yrs BP)		First dinosaurs. Extensive deserts exist in continental interiors. Climate warm. Shallow seas limited in distribution.
Phanerozoic	***Paleozoic***	***Permian*** (245-286 million yrs BP)		Permian Mass Extinction. Reptiles become more diverse. Climate cold at beginning of the Permian then warms. Average elevation of landmasses at their highest shallow seas less extensive.
		Pennsylvanian (286-320 million yrs BP)		First reptiles appear. Winged insects evolve. Occasional glaciation in Southern Hemisphere.

	Mississippian (320-360 million yrs BP)	Primitive ferns and insects evolve. Forests appear and become dominant. Mountain building producing arid habitats in the interior of some continents.
	Devonian (360-408 million yrs BP)	First amphibians and trees appear. Appalachian Mountains form. Extinction of primitive vascular plants. Landmasses generally increasing in altitude. Climate cooling.
	Silurian (408-438 million yrs BP)	Major extinction event occurs. First land plants and insects. Continents are generally flat. Tectonic uplift begins.
	Ordovician (438-505 million yrs BP)	First fish and fungi. Greatest extent of shallow seas. Climate becoming warmer.
	Cambrian (505-551 million yrs BP)	Invertebrates become common. Fossilization of the Burgess Shale. Large areas of shallow seas near the equator. Climate was warm.
Proterozoic (551-2500 million yrs BP)		Eukaryotic cell organisms develop. First multicellular organisms. Changes in the lithosphere created major land masses and extensive shallow seas.
Archean (2500-3800 million yrs BP)	**Also known as *Precambrian***	Slow development of the lithosphere, hydrosphere, and atmosphere. First single-celled prokaryotic organisms.
Hadean (3800-4600 million yrs BP)		Earth's oldest rocks come from the end of this Eon.

(c) Concept of Uniformitarianism

{PRIVATE**}*Uniformitarianism*** is one of the most important unifying concepts in the geosciences. This concept developed in the late 1700s, suggests that **catastrophic** processes were not responsible for the landforms that existed on the Earth's surface. This idea was diametrically opposed to the ideas of that time period which were based on a biblical interpretation of the history of the Earth. Instead, the theory of **uniformitarianism** suggested that the landscape developed over long periods of time through a variety of slow geologic and geomorphic processes.

The term **uniformitarianism** was first used in 1832 by **William Whewell**, a University of Cambridge scholar, to present an alternative explanation for the origin of the Earth. The prevailing view at that time was that the Earth was created through supernatural means and had been affected by a series of catastrophic events such as the biblical Flood. This theory is called **catastrophism**.

The ideas behind **uniformitarianism** originated with the work of Scottish geologist **James Hutton**. In 1785, Hutton presented at the meetings of the Royal Society of Edinburgh that the Earth had a long history and that this history could be interpreted in terms of processes currently observed. For example, he suggested that deep soil profiles were formed by the weathering of bedrock over thousands of years. He also suggested that supernatural theories were not needed to explain the geologic history of the Earth.

Figure 10c-1: James Hutton, 1726-1797.

Hutton's ideas did not gain major support of the scientific community until the work of **Sir Charles Lyell**. In the three volume publication Principles of Geology (1830-1833), **Lyell** presented a variety of geologic evidence from England, France, Italy, and Spain to prove Hutton's ideas correct and to reject the theory of **catastrophism**.

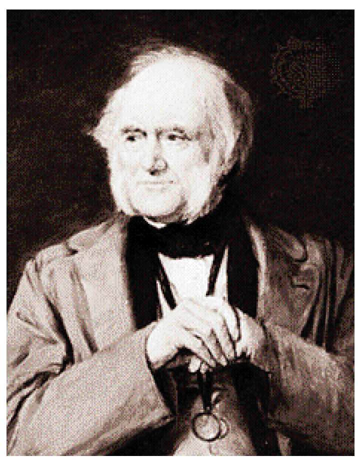


Figure 10c-2: Sir Charles Lyell, 1797-1875.

The theory of **uniformitarianism** was also important in shaping the development of ideas in other disciplines. The work of **Charles Darwin** and **Alfred Wallace** on the origin of the Earth's species extended the ideas of **uniformitarianism** into the biological sciences. The theory of ***evolution*** is based on the principle that the diversity seen in the Earth's species can be explained by the uniform modification of genetic traits over long periods of time.

Thus, uniformitarianism suggests that the continuing uniformity of existing processes should be used as the framework for understanding the geomorphic and geologic history of the Earth. Today, most theories of landscape evolution use the concept of uniformitarianism to describe how the various landforms of the Earth came to be.

(d) Composition of Rocks

{PRIVATE}A ***rock*** can be defined as a solid substance that occurs naturally because of the effects of three basic geological processes: ***magma*** solidification; ***sedimentation*** of ***weathered*** rock debris; and ***metamorphism***. As a result of these processes, three main types of rock occur:

Igneous Rocks - produced by solidification of molten magma from the mantle. Magma that solidifies at the Earth's surface conceives ***extrusive*** or ***volcanic igneous*** rocks. When magma cools and solidifies beneath the surface of the Earth ***intrusive*** or ***plutonic igneous*** rocks are formed.

Sedimentary Rocks - formed by burial, compression, and chemical modification of deposited weathered rock debris or sediments at the Earth's surface.

Metamorphic Rocks - created when existing rock is chemically or physically modified by intense heat or pressure.

Most ***rocks*** are composed of ***minerals***. **Minerals** are defined by geologists as naturally occurring ***inorganic*** solids that have a crystalline structure and a distinct chemical composition. Of course, the minerals found in the Earth's rocks are produced by a variety of different arrangements of chemical ***elements***. A list of the eight most common elements making up the minerals found in the Earth's rocks is described in **Table 10d-1**.

Table 10d-1: Common elements found in the Earth's rocks.

{PRIVATE}**Element**	**Chemical Symbol**	**Percent Weight in Earth's Crust**
Oxygen	O	46.60
Silicon	Si	27.72
Aluminum	Al	8.13
Iron	Fe	5.00
Calcium	Ca	3.63
Sodium	Na	2.83
Potassium	K	2.59
Magnesium	Mg	2.09

Over 2000 ***minerals*** have been identified by earth scientists. **Table 10d-2** describes some of the important minerals, their chemical composition, and classifies them in one of nine groups. The **Elements Group** includes over one hundred known minerals. Many of the minerals in this class are composed of only one element. Geologists sometimes subdivide this group into metal and nonmetal categories. **Gold**, **silver**, and **copper** are examples of metals. The elements sulfur and carbon produce the minerals **sulfur**, **diamonds**, and **graphite** which are nonmetallic.

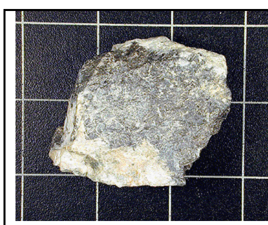


Figure 10d-1: Silver.

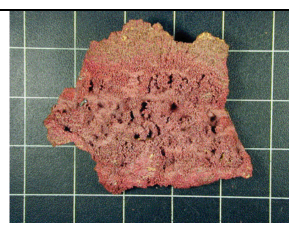


Figure 10d-2: Copper.

Figure 10d-3: Graphite.

The **Sulfide Group** are an economically important class of minerals. Many of these minerals consist of metallic elements in chemical combination with the element sulfur. Most ores of important metals such as mercury (**cinnabar** - HgS), iron (**pyrite** - FeS_2), and lead (**galena** - PbS) are extracted from sulfides. Many of the sulfide minerals are recognized by their metallic luster.

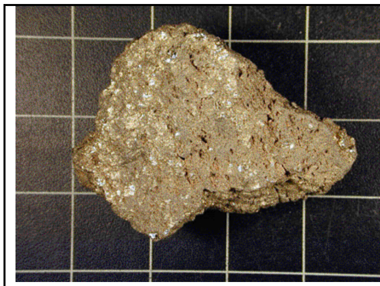


Figure 10d-4: Pyrite.

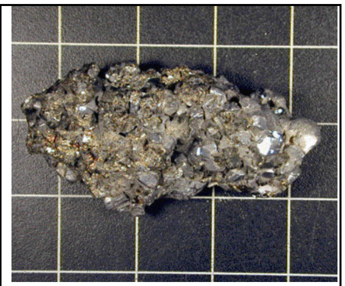


Figure 10d-5: Galena.

The **Halides** are a group of minerals whose principle chemical constituents are fluorine, chlorine, iodine, and bromine. Many of them are very soluble in water. Halides also tend to have a highly ordered molecular structure and a high degree of symmetry. The most well-known mineral of this group is **halite** (NaCl) or rock salt.

Figure 10d-6: Halite or rock salt.

The **Oxides** are a group of minerals that are compounds of one or more metallic elements combined with oxygen, water, or hydroxyl (OH). The minerals in this mineral group show the greatest variations of physical properties. Some are hard, others soft. Some have a metallic luster, some are clear and transparent. Some representative oxide minerals include **corundum**, **cuprite**, and **hematite**.

Figure 10d-7: Corundum

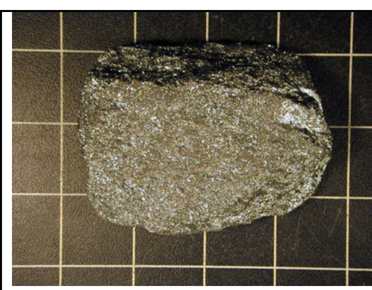


Figure 10d-8: Hematite

The **Carbonates Group** consists of minerals which contain one or more metallic elements chemically associated with the compound CO_4. Most carbonates are lightly colored and transparent when relatively pure. All carbonates are soft and brittle. Carbonates also effervesce when exposed to warm hydrochloric acid. Most geologists considered the **Nitrates** and **Borates** as subcategories of the carbonates. Some common carbonate minerals include **calcite**, **dolomite**, and **malachite**.

Figure 10d-9: Calcite.

Figure 10d-10: Dolomite.

The **Sulfates** are a mineral group that contain one or more metallic element in combination with the sulfate compound SO_4. All sulfates are transparent to translucent and soft. Most are heavy and some are soluble in water. Rarer sulfates exist containing substitutions for the sulfate compound. For example, in the **chromates** SO_4 is replaced by the compound CrO_4. Two common sulfates are **anhydrite** and **gypsum**.

Figure 10d-11: Gypsum.

The **Phosphates** are a group of minerals of one or more metallic elements chemically associated with the phosphate compound PO_4. The phosphates are often classified together with the arsenate, vanadate, tungstate, and molybdate minerals. One common phosphate mineral is **apatite**. Most phosphates are heavy but soft. They are usually brittle and occur in small crystals or compact aggregates.

The **Silicates** are by far the largest group of minerals. Chemically, these minerals contain varying amounts of silicon and oxygen. It is easy to distinguish silicate minerals from other groups, but difficult to identify individual minerals within this group. None are completely opaque. Most are light in weight. The construction component of all silicates is the ***tetrahedron***. A tetrahedon is a chemical structure where a silicon ***atom*** is joined by four oxygen atoms (SiO_4). Some representative minerals include **albite**, **augite**, **beryl**, **biotite**, **hornblende**, **microcline**, **muscovite**, **olivine**, **othoclase**, and **quartz**.

Figure 10d-12: Albite.

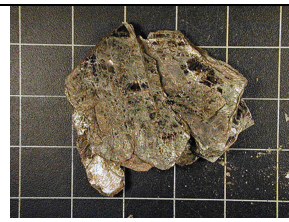


Figure 10d-12: Biotite.

Figure 10d-13: Hornblende.

Figure 10d-14: Olivine.

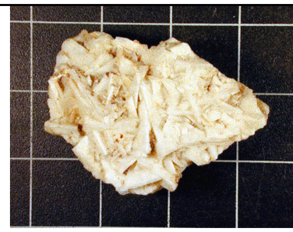


Figure 10d-15: Orthoclase.

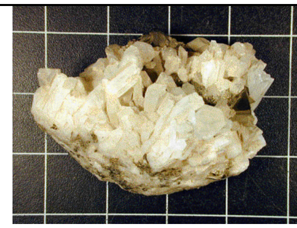


Figure 10d-16: Quartz.

The **Organic** minerals are a rare group of minerals chemically containing ***hydrocarbons***. Most geologists do not classify these substances as true minerals. Note that our original definition of a mineral excludes organic substances. However, some organic substances that are found naturally on the Earth that exist as crystals that resemble and act like true minerals. These substances are called organic minerals. **Amber** is a good example of an organic mineral.

Table 10d-2: Classification of some of the important minerals found in rocks.

{PRIVATE}*Group*	*Typical Minerals (and information link)*	*Chemistry*
Elements	*Gold*	Au
	Silver	Ag
	Copper	Cu
	Carbon (*Diamond* and *Graphite*)	C
	Sulfur	S
Sulfides	*Cinnabar*	HgS
	Galena	PBS
	Pyrite	FeS_2
Halides	*Fluorite*	CaF_2
	Halite	NaCl
Oxides	*Corundum*	Al_2O_3
	Cuprite	Cu_2O
	Hematite	Fe_2O_3
Carbonates (Nitrates and Borates)	*Calcite*	$CaCO_3$
	Dolomite	$CaMg(CO_3)_2$
	Malachite	$Cu_2(CO_3)(OH)_2$
Sulfates	*Anhydrite*	$CaSO_4$
	Gypsum	$CaSO_4$-2(H_2O)
Phosphates (Arsenates, Vanadates, Tungstates, and Molybdates)	*Apatite*	$Ca_5(F,Cl,OH)(PO_4)$
Silicates	*Albite*	$NaAlSi_3O_8$
	Augite	$(Ca, Na)(Mg, Fe, Al)(Al, Si)_2 O_6$
	Beryl	$Be_3Al_2(SiO_3)_6$
	Biotite	$K (FE, Mg)_3 AlSi_3 O_{10} (F, OH)_2$
	Hornblende	$Ca_2(Mg, Fe, Al)_5 (Al, Si)_8O_{22}(OH)_2$
	Microcline	$KAlSi_3O_8$
	Muscovite	$KAl_2(AlSi_3O_{10})(F, OH)_2$
	Olivine	$(Mg, Fe)_2SiO_4$
	Orthoclase	$KAlSi_3O_8$
	Quartz	SiO_2
Organics	*Amber*	$C_{10}H_{16}O$

(e) Characteristics of Ig neous Rocks

{PRIVATE}Introduction

As described in some of the previous topics, ***igneous rocks*** are produced by the crystallization and solidification of molten magma. Magma forms when rock is heated to high temperatures (between 625 and 1200 degrees Celsius) beneath the Earth's surface. The exact temperature needed to melt rock is controlled by several factors. Chemistry of the rock material, pressure, presence of gases (like water vapor) all influence when melting occurs. Most of the heat required to melt rock into magma comes from the Earth's central internal region known as the ***core***. Scientists estimate that the temperature of the Earth's core is about 5000 degrees Celsius. Heat moves from the Earth's core towards the solid outer crust by ***convection*** and ***conduction***. Convection moves hot ***plumes of magma*** vertically from the ***lower mantle*** to the ***upper mantle***. Some of these plumes melt through the Earth's solid lithosphere and can produce ***intrusive igneous*** features and ***extrusive igneous*** features on the surface. Heat can also be generated in the lower ***lithosphere*** through ***friction***. The ***tectonic*** movement of ***subducted crustal plates*** can generate enough heat (and pressure) to melt rock. This fact explains the presence of ***volcanoes*** along the margin of some continental plates.

Types of Igneous Rocks

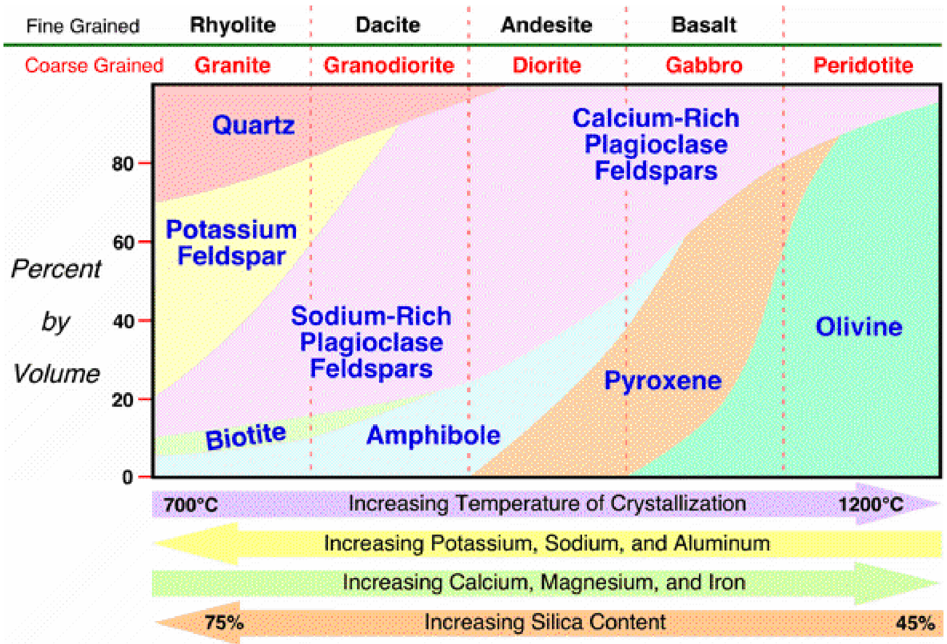
The type of igneous rocks that form from magma is a function of three factors: the chemical composition of the magma; temperature of solidification; and the rate of cooling which influences the crystallization process. Magma can vary chemically in its composition. For example, the amount of ***silica*** (SiO_2) found in magma can vary from 75 % to less than 45 %. The temperature of cooling determines which types of minerals are found dominating the rock's composition. Rocks that begin their cooling at low temperatures tend to be rich in minerals composed of silicon, potassium, and aluminum. High temperature igneous rocks are dominated by minerals with higher quantities of calcium, sodium, iron, and magnesium. The rate of cooling is important in crystal development. Igneous rocks that form through a gradual cooling process tend to have large crystals. Relatively fast cooling of magma produces small crystals. Volcanic magma that cools very quickly on the Earth's surface can produce ***obsidian*** (see image ***link***) glass which contains no crystalline structures.

Geologists have classified the chemistry of igneous rocks into four basic types: ***felsic***, intermediate, ***mafic***, and ***ultramafic***. Igneous rocks derived from ***felsic magma*** contain relatively high quantities of sodium, aluminum, and potassium and are composed of more than 65 % silica. Rocks formed from felsic magma include ***granite*** (see image ***link***), granodiorite (see image ***link***), dacite, and ***rhyolite*** (see image ***link***). All of these rock types are light in color because of the dominance of ***quartz***, ***potassium and sodium feldspars***, and ***plagioclase feldspar*** minerals (**Figure 10e-1**). Dacite and granodiorite contain slightly more ***biotite*** and ***amphibole*** minerals than granite and rhyolite. Rhyolite and dacite are produced from continental ***lava*** flows that solidify quickly. The quick solidification causes the mineral crystals in these rocks to be fine grained. Granite and granodiorite are common intrusive igneous rocks that are restricted to the Earth's continents. Large expanses these rocks were formed during episodes of mountain building on the Earth. Because granite and granodiorite form beneath the Earth's surface their solidification is a relatively slow process. This slow solidification produces a rock with a coarse mineral grain.

Mafic magma produces igneous rocks rich in calcium, iron, and magnesium and are relatively poor in ***silica*** (silica amounts from 45 to 52 %). Some common mafic igneous rocks include fine grained ***basalt*** (see image ***link***) and coarse grained ***gabbro*** (see image ***link***). Mafic igneous rocks tend to be dark in color because they contain a large proportion of minerals rich in iron and magnesium (***pyroxene***, ***amphiboles***, and ***olivine***). Basalt is much more common than gabbro. It is found in the upper portion of the oceanic crust and also in vast continental lava flows that cover parts of Washington, Oregon, Idaho, and California. Gabbro is normally found in the lower parts of oceanic crust and sometimes in relatively small intrusive features in continental crust.

Andesite (see image ***link***) and ***diorite*** are intermediate igneous rocks that have a chemistry between mafic and felsic (silica amounts between 53 to 65 %). These rocks are composed predominantly of the minerals ***plagioclase feldspar***, ***amphibole***, and ***pyroxene***. Andesite is a common fine grained extrusive igneous rock that forms from lavas erupted by volcanoes located along continental margins. Coarse grained diorite is found in intrusive igneous bodies associated with continental crust.

Ultramafic igneous rocks contain relative low amounts of silica (< 45 %) and are dominated by the minerals ***olivine***, calcium-rich ***plagioclase feldspars***, and ***pyroxene***. ***Peridotite*** is the most common ultramafic rock found in the Earth's crust. These rocks are extremely rare at the Earth's surface.



{PRIVATE**}Figure 10e-1:** The classification of igneous rocks. This graphic model describes the difference between nine common igneous rocks based on texture of mineral grains, temperature of crystallization, relative amounts of typical rock forming elements, and relative proportions of silica and some common minerals.

Igneous Rocks and the Bowen Reaction Series

In the 1920s, N.L. Bowen created the following model to explain the origin of the various types of igneous rocks (**Figure 10e-2**). This model, known as the ***Bowen reaction series***, suggests that the type of igneous rocks that form from magma solidification depends on the temperature of crystallization and the chemical composition of the originating magma. Bowen theorized that the formation of minerals, which make up ***igneous rocks***, begins with two different chemical sequences at high temperatures that eventually merge into a single series at cooler temperatures. One sequence, the **discontinuous series**, involves the formation of chemically unique minerals at discrete temperature intervals from iron and magnesium rich ***mafic magma***. In the other sequence, known as the **continuous series**, temperature reduction causes a gradual change in the chemistry of the minerals that form calcium and sodium rich ***felsic magma***. The discontinuous series starts with the formation of rocks that are primarily composed of the mineral ***olivine***. Continued temperature decreases change the minerals dominating the composition of the rock from ***pyroxene***, to ***amphibole***, and then ***biotite***. The continuous series produces light colored rocks rich in ***plagioclase feldspar*** minerals. At high temperatures, the **plagioclase feldspar** minerals are dominated with the element calcium. With continued cooling, the calcium in these minerals is gradually replaced with sodium. The convergence of both series occurs with a continued drop in magma temperature. In the merged series, the minerals within the crystallizing rock become richer in potassium and silica and we get the formation of first **potassium feldspars** and then the mineral ***muscovite***. The last mineral to crystallize in the Bowen reaction series is ***quartz***. Quartz is a silicate mineral composed of just silicon and oxygen (SiO_2).

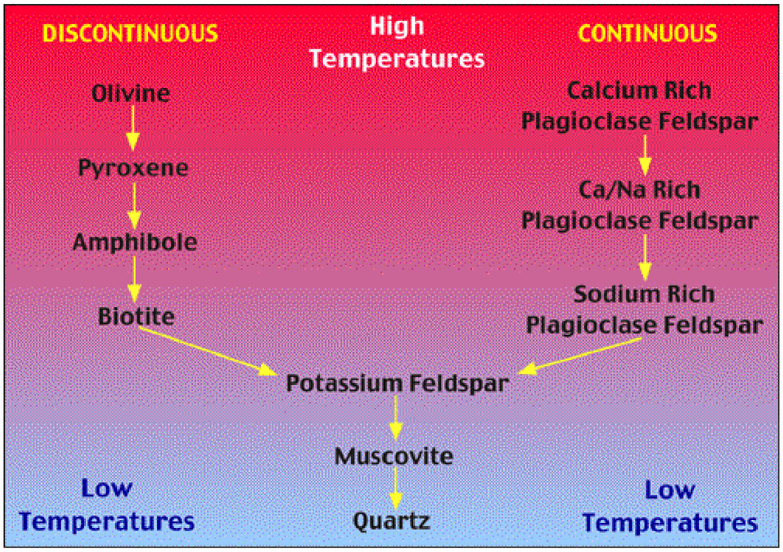


Figure 10e-2: Bowen reaction series.

(f) Characteristics of Sedimentary Rocks

{PRIVATE**}*****Sedimentary rocks*** can be categorized into three groups based on sediment type. Most sedimentary rocks are formed by the ***lithification*** of ***weathered*** rock debris that has been physically transported and deposited. During the transport process, the particles that make up these rocks often become rounded due to abrasion or can become highly sorted. Examples of this type of sedimentary rock include ***conglomerate*** and ***sandstone***. Scientists sometimes call this general group of sedimentary rocks ***clastic***. The remaining types of sedimentary rocks are created either from chemical precipitation and crystallization, or by the lithification of once living ***organic matter***. We identify these sedimentary rocks as being ***non-clastic***.

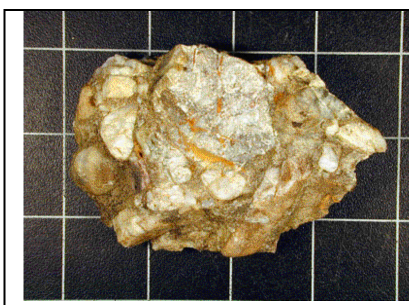


Figure 10f-1: Conglomerate.

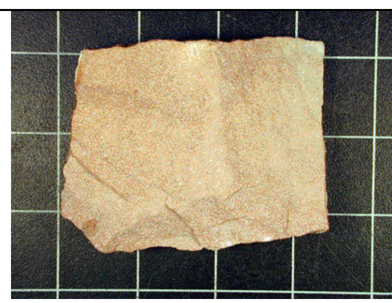


Figure 10f-2: Sandstone.

All sedimentary rocks are lithified into some collective mass. ***Lithification*** is any process that turns raw rock sediment into consolidated sedimentary rock. The process of lithification usually produces identifiable layering in these type of rocks (**Figure 10f-3**). Lithification can occur by way of:

Drying and compaction.

Oxidation of iron and aluminum.

Precipitation of calcium and silica.

Figure 10f-3: Dipping sedimentary layers of rock, Rocky Mountains, Canada.

The classification of clastic sedimentary rocks is based on the particle types found in the rock. Some types of clastic sedimentary rocks are composed of weathered rock material like ***gravel***, ***sand***, ***silt***, and ***clay***. Others can be constructed from the break up and deposition of shells, coral and other marine organisms by wave-action and ocean currents. **Table 10f-1** describes some of the main types of clastic sedimentary rocks.

Table 10f-1: Clastic sedimentary rocks.

{PRIVATE}**Name of Rock**	**Fragment Type**
Breccia (Image *Link*)	Coarse Fragments of Angular ***Gravel*** and Rocks
Conglomerate	Coarse Fragments of Rounded ***Gravel*** and Rocks
Sandstone	***Sand*** Sized Particles that are 90 % ***Quartz***
Arkose	Sandstone composed of 25 % ***Feldspar*** Grains
Shale	***Clay*** Particles
Siltstone	***Silt*** Particles
Mudstone	Mixture of ***Clay*** and ***Silt***
Limestone	Mixture of Shells, Coral, and Other Marine Skeletons

Figure 10f-4: Arkose.

Figure 10f-5: Shale.

Figure 10f-6: Siltstone.

Figure 10f-7: Limestone.

Earlier it was suggested that there were two types of non-clastic sedimentary rocks. One group forms through the chemical precipitation and crystallization of elements and compounds from solution. Elements such as calcium, sodium, potassium, and magnesium are commonly released into the environment through a variety of ***chemical weathering*** processes. These elements can then become dissolved into aqueous solutions that are often transported via ***runoff***, ***stream flow***, or ***groundwater flow***. If this solution enters a basin environment where ***evaporation*** exceeds ***precipitation*** and in-flow, sedimentary ***evaporites*** can form because of the loss of water from the solution.

The oceans are almost saturated with dissolved ***calcium carbonate***. This compound originates from the shells of a variety of marine organisms that use it for the construction of shells and other hard body parts. Because these organisms are surrounded in a solution, some of the calcium carbonate dissolves into the ocean waters. Under the right circumstances the dissolved calcium carbonate can precipitate out forming chemically created ***limestone*** deposits. The formation of ***dolomite*** involves the chemical modification of limestone deposits by a magnesium rich solution.

Figure 10f-8: Dolomite.

The following **table** describes some of the common forms of chemical precipitated sedimentary rocks.

Table 10f-2: Sedimentary rocks formed as chemical precipitates.

{PRIVATE**}Name of Rock**	**Precipitate Type**
Halite	Sodium and Chlorine
Gypsum	Calcium, Sulfur, and Oxygen
Silcretes	Silica
Ferricretes	Iron
Limestone	Calcium Carbonate
Dolomite	Calcium Magnesium Carbonate

Figure 10f-9: Halite.

Figure 10f-10: Gypsum.

Several types of sedimentary rocks are formed from the lithification of once living organisms. ***Limestone*** deposits can be formed by the direct lithification of coral reefs, marine organism shells, or marine organism skeletons. ***Chalk*** is a particular variety of limestone that is composed of the skeletons of marine microorganisms like ***forminifera***. ***Coal*** and ***lignite*** are the lithified remains of plants.

Figure 10f-11: Chalk.

Figure 10f-12: Coal.

(g) Characteristics of Metamorphic Rocks

{PRIVATE**}Introduction**

Metamorphism involves the alteration of existing rocks by either excessive heat and pressure, or through the chemical action of fluids. This alteration can cause chemical changes or structural modification to the minerals making up the rock. Structural modification may involve the simple reorganization of minerals into layers or the aggregation of minerals into specific areas within the rock.

Much of the Earth's ***continental crust*** is composed of metamorphic and ***igneous*** rocks. Together, these two rock types form the base material at the core of the Earth's major continental masses. Overlying this core are often thick layers of ***sedimentary rocks***. In some regions, this base rock is exposed to the atmosphere and is known as ***shields***. On the ***Canadian Shield*** we can find some of the oldest rocks found on the planet (3.96 billion years old). These very old rocks are primarily metamorphic. Metamorphic rocks also are the rock type found at the core of the world's various mountain ranges.

Heat and Metamorphism

Heat is an important agent in the metamorphic modification of rock. Rocks begin to change chemically at temperatures above 200 degrees Celsius. At these temperatures, the crystalline structure of the ***minerals*** in the rock are broken down and transformed using different combinations of the available ***elements*** and ***compounds***. As a result, new minerals are created. The metamorphic process stops when the temperatures become high enough (600 to 1200 degrees Celsius) to cause complete melting of the rock. If rocks are heated to the point where they become ***magma***, the magma when cooled creates new igneous rocks. Thus, metamorphism only refers to the alteration of rock that takes place before complete melting occurs.

Heat can be applied to rock through two processes: ***tectonic subduction*** and the ***intrusion*** of ***magma***. Some rocks that are formed at the surface are subsequently transported deep into the ***crust*** and the ***upper mantle*** at ***tectonic subduction zones***. Temperatures beneath the Earth's surface increase with depth at a rate of about 25 degrees Celsius per kilometer. Scientists estimate that the temperature at the base of the crust is about 800 to 1200 degrees Celsius. This heat is generated from the decay of radioactive materials in the ***mantle*** and the ***core***.

Magma can sometimes migrate up through the crust forming an ***igneous intrusion***. This is especially true along continental boundaries, like the western side of North America, where subduction is taking place. Metamorphism takes place in the rock surrounding the magma body because of heat dissipation. Because of the nature of the dissipation process, the level of metamorphic alteration in the influenced rock decreases with distance from the igneous intrusion.

Pressure and Metamorphism

Rocks that buried are subjected to pressure because of the weight of overlying materials. Pressure can also be exerted on rocks due the forces involved in a variety of tectonic processes. The most obvious effect of pressure on rocks is the reorientation of ***mineral*** crystals. Under extreme levels of pressure rocks become plastic creating flow structures in their crystalline structure. Pressure almost never acts in isolation as temperatures do get higher with increasing depth below the Earth's surface.

Chemical Action of Fluids

Water and carbon dioxide are often found in small amounts in the perimeter between mineral crystals or in the pore spaces of rocks. When mixed, the resulting fluid enhances metamorphism by dissolving ions and by causing chemical reactions. Usually, the end product of this process is the creation of new minerals by the substitution, removal, or addition of chemical ions. Sometimes fluids can also permeate into rock from adjacent magma.

Types of Metamorphism

Geologists suggest that metamorphism can occur by way of the following three processes.

Thermal metamorphism involves the heating and structural and chemical alteration of rocks through processes associated with ***plate tectonics***. This type of metamorphism has two sub-categories:

Regional metamorphism is the large scale heating and modification of existing rock through the creation of ***plutons*** at tectonic zones of ***subduction***. It involves large areas and large volumes of rock.

Contact metamorphism is the small scale heating and alteration of rock by way of a localized igneous intrusion (for example, volcanic ***dykes*** or ***sills***).

Dynamic metamorphism causes only the structural alteration of rock through pressure. The minerals in the altered rocks do not change chemically. The extreme pressures associated with mountain building can cause this type of metamorphism.

Metasomatic metamorphism involves the chemical replacement of elements in rock minerals when gases and liquids permeate into bedrock.

Common Metamorphic Rocks

Examples of metamorphic rock types include: ***Slate*** is a fine grained metamorphic rock. It is created by minor metamorphism of ***shale*** or ***mudstone***. This rock is characterized by the ***foliation*** (**Figure 10g-6**) of its mineral grains which causes it to have ***cleavage*** that is parallel.

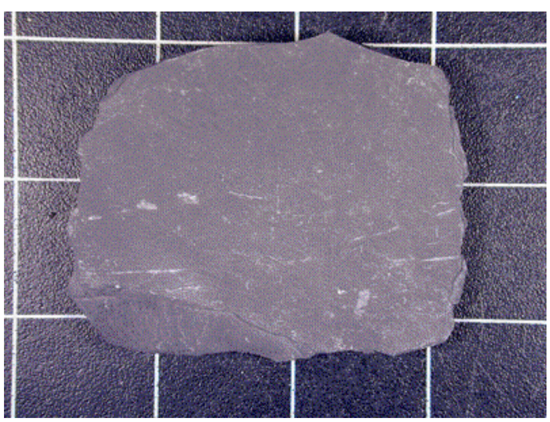


Figure 10g-1: Slate.

Schist is a medium to coarse grained foliated rock. Foliation is the result of the rearrangement of mica, chlorite, talc, and hematite mineral grains into parallel structures. When compared to slate, schists result from more intense metamorphism.

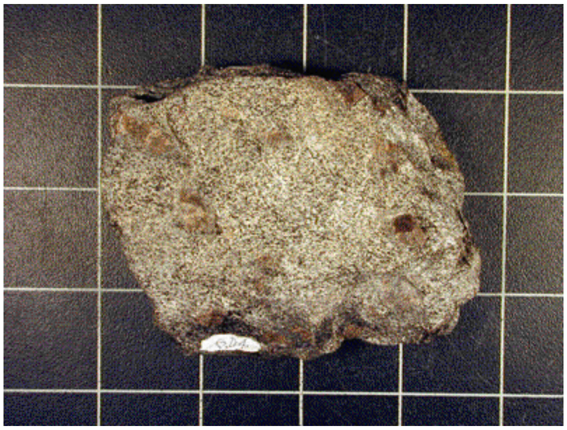


Figure 10g-2: Schist.

Gneiss is a coarse grained metamorphized igneous rock. In this rock, you get the recrystallization and foliation of ***quartz***, ***feldspars***, ***micas***, and ***amphiboles*** into alternating light and dark colored bands.

Figure 10g-3: Gneiss.

Marble is a nonfoliated metamorphized ***limestone*** or ***dolomite***).

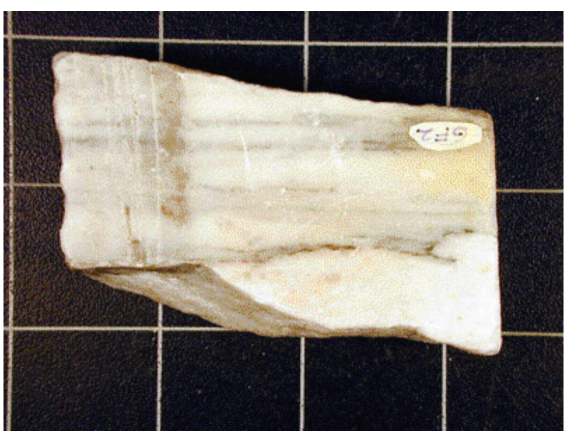
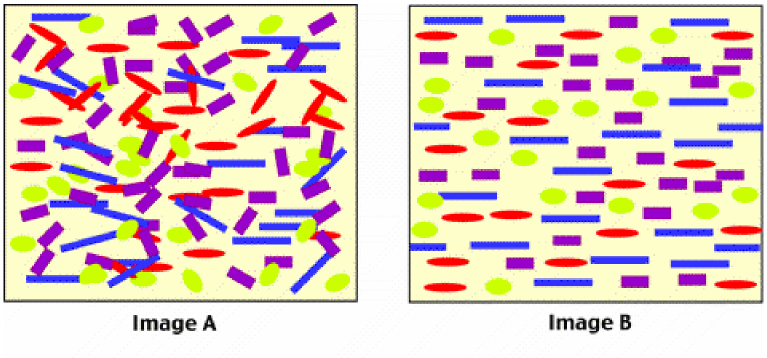


Figure 10g-4: Marble.

Quartzite forms from the recrystallization of silica found in ***sandstone***.

Figure 10g-5: Quartzite.



{PRIVATE**}Figure 10g-6:** The mineral grains in rocks subjected to extreme pressure often rearrange themselves in a parallel fashion, creating a ***foliated*** texture (**Image A** - before metamorphism; **Image B** - after metamorphism).

(h) Structure of the Earth

{PRIVATE**}**The Earth is an oblate spheroid. It is composed of a number of different layers as determined by deep drilling and ***seismic*** evidence (**Figure 10h-1**). These layers are:

The ***core*** which is approximately 7000 kilometers in diameter (3500 kilometers in radius) and is located at the Earth's center.

The ***mantle*** which surrounds the core and has a thickness of 2900 kilometers.

The ***crust*** floats on top of the mantle. It is composed of ***basalt*** rich oceanic crust and ***granitic*** rich continental crust.

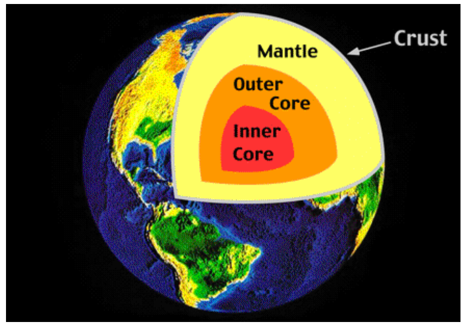


Figure 10h-1: Layers beneath the Earth's surface.

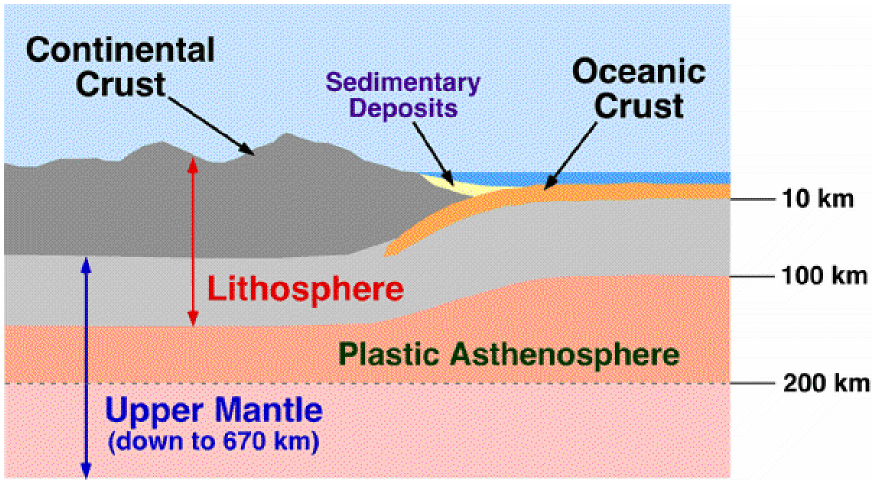
The core is a layer rich in iron and nickel that is composed of two layers: the **inner** and **outer cores**. The ***inner core*** is theorized to be solid with a density of about 13 grams per cubic centimeter and a radius of about 1220 kilometers. The ***outer core*** is liquid and has a density of about 11 grams per cubic centimeter. It surrounds the inner core and has an average thickness of about 2250 kilometers.

The ***mantle*** is almost 2900 kilometers thick and comprises about 83 % of the Earth's volume. It is composed of several different layers. The ***upper mantle*** exists from the base of the crust downward to a depth of about 670 kilometers. This region of the Earth's interior is thought to be composed of ***peridotite***, an ultramafic rock made up of the minerals ***olivine*** and ***pyroxene***. The top layer of the upper mantle, 100 to 200 kilometers below surface, is called the ***asthenosphere***. Scientific studies suggest that this layer has physical properties that are different from the rest of the upper mantle. The rocks in this upper portion of the mantle are more rigid and brittle because of cooler temperatures and lower pressures. Below the upper mantle is the ***lower mantle*** that extends from 670 to 2900 kilometers below the Earth's surface. This layer is hot and plastic. The higher pressure in this layer causes the formation of minerals that are different from those of the upper mantle.

The ***lithosphere*** is a layer that includes the crust and the upper most portion of the asthenosphere (**Figure 10h-2**). This layer is about 100 kilometers thick and has the ability to glide over the rest of the upper mantle. Because of increasing temperature and pressure, deeper portions of the lithosphere are capable of plastic flow over geologic time. The lithosphere is also the zone of ***earthquakes***, ***mountain building***, ***volcanoes***, and ***continental drift***.

The topmost part of the lithosphere consists of crust. This material is cool, rigid, and brittle. Two types of crust can be identified: ***oceanic crust*** and ***continental crust*** (**Figure 10h-2**). Both of these types of crust are less dense than the rock found in the underlying upper mantle layer. Ocean crust is thin and measures between 5 to 10 kilometers thick. It is also composed of ***basalt*** and has a density of about 3.0 grams per cubic centimeter.

The continental crust is 20 to 70 kilometers thick and composed mainly of lighter ***granite*** (**Figure 10h-2**). The density of continental crust is about 2.7 grams per cubic centimeter. It is thinnest in areas like the ***Rift Valleys*** of East Africa and in an area known as the Basin and Range Province in the western United States (centered in Nevada this area is about 1500 kilometers wide and runs about 4000 kilometers North/South). Continental crust is thinkest beneath mountain ranges and extends into the mantle. Both of these crust types are composed of numerous ***tectonic plates*** that float on top of the mantle. Convection currents within the mantle cause these plates to move slowly across the asthenosphere.

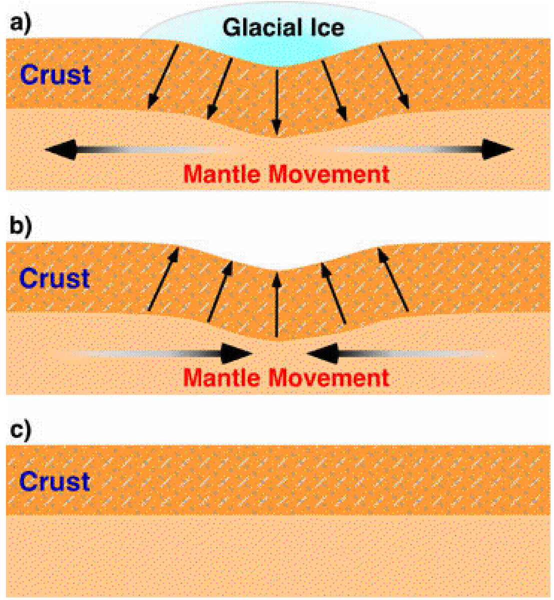


{PRIVATE**}Figure 10h-2:** Structure of the Earth's crust and top most layer of the upper mantle. The lithosphere consists of the oceanic crust, continental crust, and uppermost mantle. Beneath the lithosphere is the asthenosphere. This layer, which is also part of the upper mantle, extends to a depth of about 200 kilometers. ***Sedimentary*** deposits are commonly found at the boundaries between the continental and oceanic crust.

Isostacy

One interesting property of the continental and oceanic crust is that these ***tectonic plates*** have the ability to rise and sink. This phenomenon, known as ***isostacy***, occurs because the crust floats on top of the mantle like ice cubes in water. When the Earth's crust gains weight due to mountain building or glaciation, it deforms and sinks deeper into the mantle (**Figure 10h-3**). If the weight is removed, the crust becomes more buoyant and floats higher in the mantle.

This process explains recent changes in the height of ***sea-level*** in coastal areas of eastern and northern Canada and Scandinavia. Some locations in these regions of the world have seen sea-level rise by as much as one meter over the last one hundred years. This rise is caused by ***isostatic rebound***. Both of these areas where covered by massive glacial ***ice sheets*** about 10,000 years ago. The weight of the ice sheets pushed the crust deeper into the mantle. Now that the ice is gone, these areas are slowly increasing in height to some new ***equilibrium*** level.



{PRIVATE**}Figure 10h-3:** The addition of glacial ice on the Earth's surface causes the crust to deform and sink (**a**).When the ice melts, isostatic rebound occurs and the crust rises to its former position before glaciation (**b** and **c**). A similar process occurs with mo untain building and mountain erosion (see **topic *10l***).

(i) Plate Tectonics

{PRIVATE**}**In the 19th and early 20th centuries, several scientists suggested that the continental masses had the ability to move across the Earth's surface. These early theories of ***continental drift*** were based on the following evidence:

Locations of fossil occurrences suggested that some of the continental masses may have been connected in the geological past.

Paleoclimatic evidence indicates that now tropical regions on some continents had polar climates in the past. This may indicate that these regions were located at different latitudes.

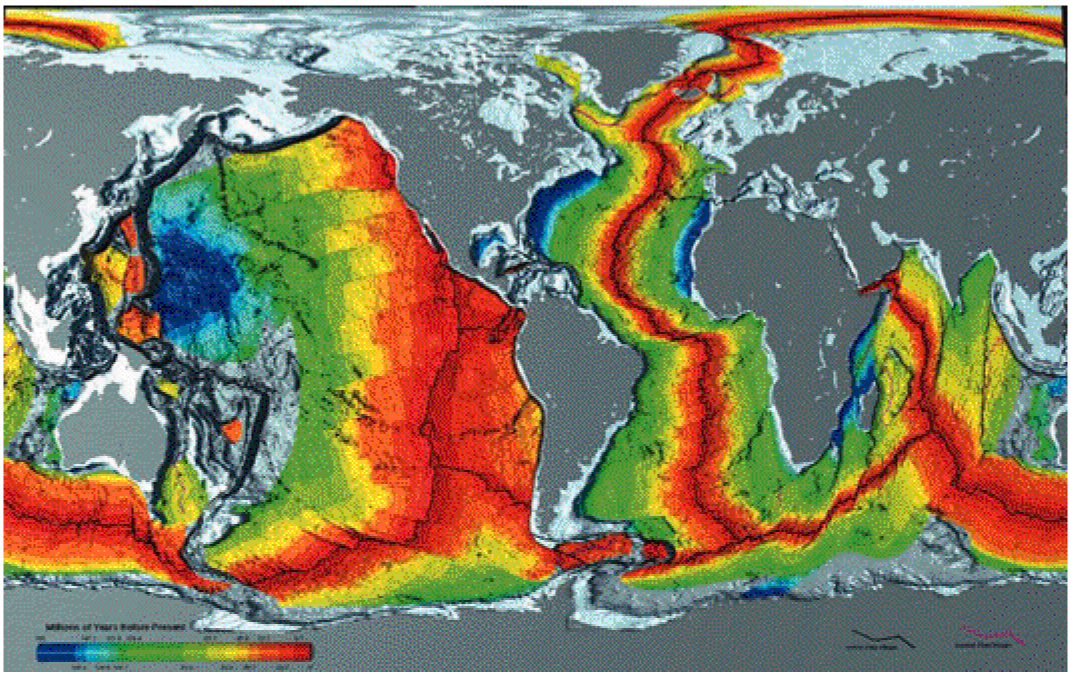
Some continents seem to fit together like a jigsaw puzzle.

Some geologic deposits of rocks on the East coast of North and South America are similar to deposits found on the West coast of Africa and Europe.

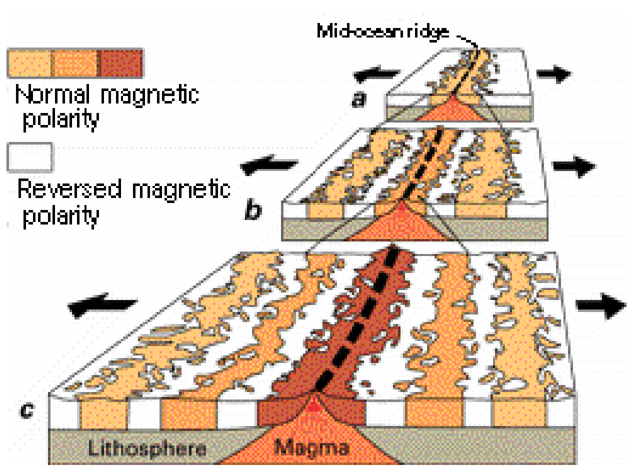
During the first 30 years of this century the theory of ***continental drift*** was actively debated among geo-scientists. However, during the following 30 year period, debate on this theory waned because of the inability of scientists to propose a mechanism to cause the movement of the continental masses.

In the 1960s, the theory was resurrected with the discovery of alternating patterns of rock magnetism in surface sea-floor rocks. Scientists had previously discovered that the magnetic orientation of certain crystals in rocks varies from normal to reversed polarity depending on the date that the rock was formed and solidified. It was also discovered that these magnetic reversals were common and occurred on a regular basis. The polarity patterns found in the rocks at the ocean floor seemed to mirror themselves either side of the ***mid-oceanic ridge*** found at the centers of the ***ocean basins***. Further, geologic dating of the rocks indicated the age of the sea-floor rocks increase as one moved away from the mid-oceanic ridge (see ***Figure 10i-1***). Based on this information, scientists developed the theory of ***sea-floor spreading*** which suggested that volcanic ***rift zones*** at the mid-oceanic ridge represent areas of crustal creation. The following diagram illustrates the process of crustal creation and the magnetic striping process. In **Figure 10i-2**, **illustrations "a"** to **"c"** represent a sequence in time from the past to the present. In **illustration a**, rocks of normal polarity are being deposited at the rift zone located along the mid-oceanic ridge. As new rock is created, older rock is pushed away from the ridge. The reversed polarity rock shown in this diagram was created before the current normal polarity layer. **Illustration b** shows the process some time later. In this diagram, we now have four layers of rock with alternating polarity. By the third **illustration**, **sea-floor spreading** and changes in magnetic polarity have created six recognizable layers of rock either side of the **rift zone**.

Age of the Oceanic Crust



{PRIVATE**}Figure 10i-1:** In the following image, the age of the ***oceanic crust*** is illustrated by color. The gradation from red to blue indicates increasing age. Blue represents crust created some 180 million years ago. Red indicates oceanic crust created quite recently on the geologic time scale. Center black lines delineate the ***mid-oceanic ridge*** **volcanic** ***rift zones***. (**Source:** *National Geophysical Data Center*, National Oceanic and Atmospheric Administration).



{PRIVATE}**Figure 10i-2:** Creation of oceanic crust on the ocean floor. (**Source:** U.S. Geological Survey).

The theory of **sea-floor spreading** started a revolution in the Earth Sciences. Subsequent research discovered that the Earth's surface was composed of a number of ***oceanic*** and ***continental plates*** that float on top of the ***asthenosphere*** (see ***Figure 10i-3***). Other research suggested that convection currents within the Earth's mantle were responsible for the creation of ***oceanic crust*** and the drifting of the continents (**Figure 10i-4**). In this diagram, it is theorized that convection currents within the Earth's ***mantle*** cause the creation of new **oceanic crust** at the **mid-oceanic ridges**. Oceanic crust is destroyed at areas where this crust type becomes subducted under lighter ***continental crust***. This process also creates the deep ***oceanic trenches***.

Earth's Tectonic Plates

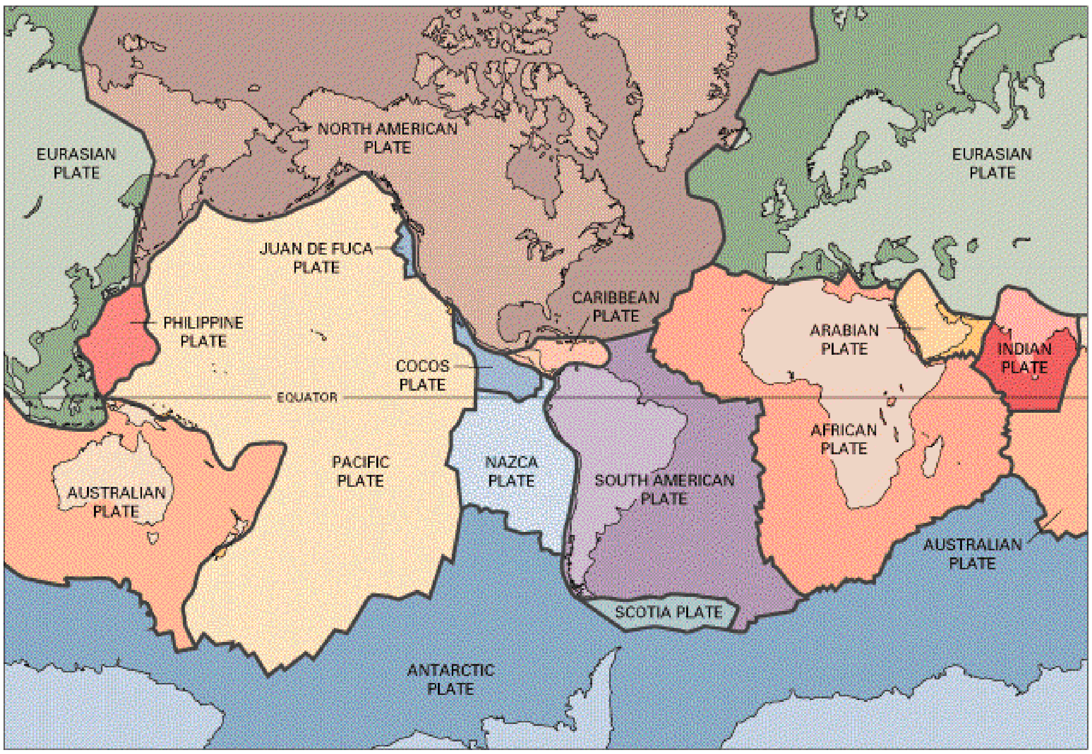
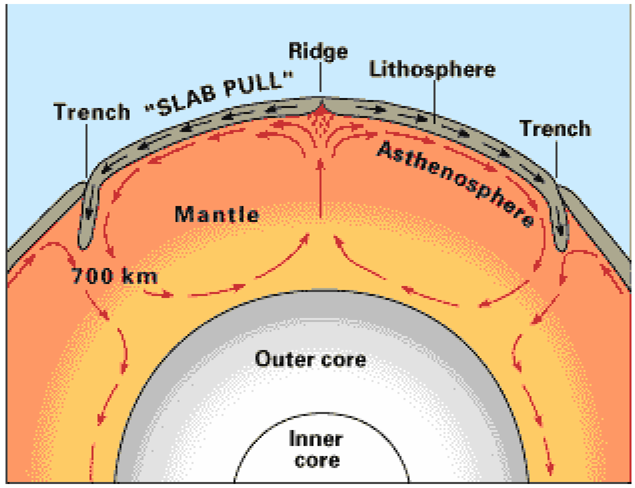
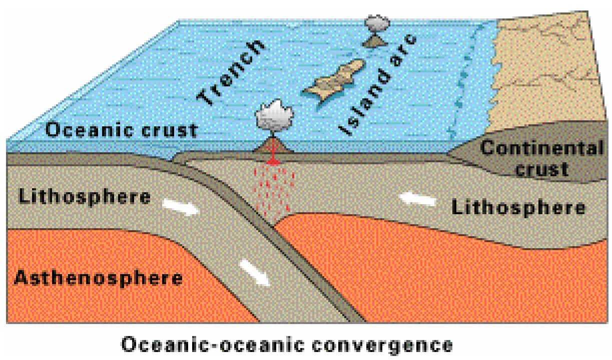


Figure 10i-3: Earth's tectonic plates. (**Source:** U.S. Geological Survey).



{PRIVATE**}Figure 10i-4:** Convection currents in the Earth's mantle and their role in oceanic crust formation and destruction. (**Source:** U.S. Geological Survey).

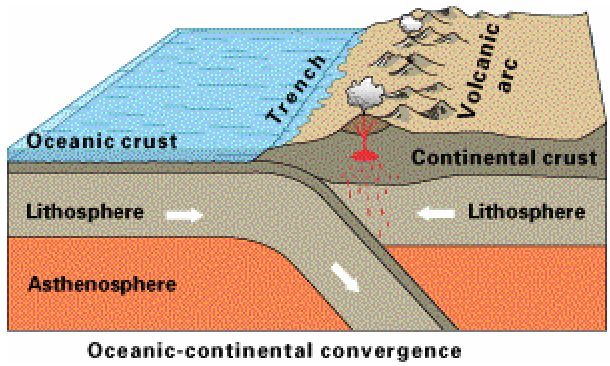
The theory of **plate tectonics** offered new and more scientifically sound explanations for a number of observed geologic phenomena. For example, the following **diagrams** illustrate the three types of **plate convergence** and describe some of the geologic repercussions of these processes. The first diagram models the tectonic convergence of two **oceanic plates** (**Figure 10i-5**).



{PRIVATE**}Figure 10i-5:** Collision of two oceanic plates. (**Source:** U.S. Geological Survey).

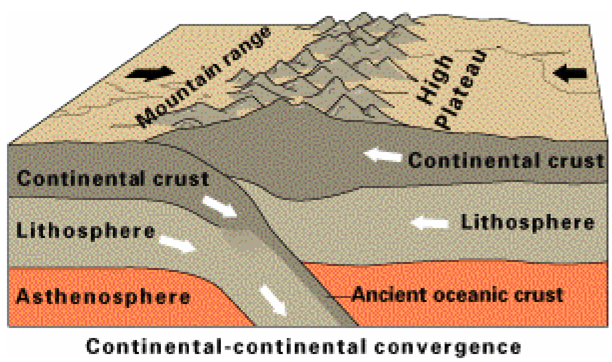
In this type of a collision, one of the plates is subducted under the other creating a deep ***oceanic trench***. The **Marianas trench** in the Pacific ocean is created by the collision of the fast-moving Pacific Plate against the slower moving Philippine Plate. Convergence of two oceanic plates also creates chains of volcanic islands called ***island arcs***. Island arcs are created by the friction of subduction which creates hot plumes of magma at the interface of the two plates. These hot plumes of magma then rise to the Earth's surface to form ***volcanoes***. Another phenomena associated with collision and subduction of the plates is ***earthquakes***. The gliding of one plate under the other is not smooth but jerky producing ***seismic waves***.

The next diagram shows the collision of an **oceanic** and a **continental plate** (**Figure 10i-6**). In this illustration, the oceanic plate subducts under the lighter continental plate. Once again we get the formation of a **trench**, **volcanoes**, and **earthquakes**. Collision causes sediment deposited on the ocean floor to be piled up at the continental plate boundary. The creation of hot ***magma plumes*** also causes the **continental crust** to deform producing **mountains**.



{PRIVATE**}Figure 10i-6:** Collision of a oceanic plate with a continental plate. (**Source:** U.S. Geological Survey).

In the final illustration two **continental plates** collide (**Figure 10i-7**). Once again one of the crustal plates is subducted under the other producing **earthquakes**. A **mountain range** is produced at the plate boundaries because of the deformation of rocks. Some of the rocks in the mountain range may be ***sedimentary*** and may have been set down in an ocean environment that existed between the two continental crusts prior to collision.



{PRIVATE**}Figure 10i-7:** Collision of two continental plates. (**Source:** U.S. Geological Survey).

In summary, modern **plate tectonic** theory states that the surface crust of the Earth is composed of many independent segments called plates. These plates have the ability to move horizontally by gliding over the plastic **asthenosphere** (see ***Figure 10i-8***). In some cases, plates can collide with each other at the plate boundaries causing subduction and the production of **earthquakes**, **volcanoes**, **mountain building**, and **oceanic trenches**. At other plate boundaries, plates may move away from each other because of **sea-floor spreading** or horizontally move past one another creating ***transform faults*** and **earthquakes**.

(j) Crustal Formation Processes

{PRIVATE**}**Studies of ***seismic waves*** have discovered that the Earth's crust consists of two basic types. Beneath the oceans we find a crust that is on average 7 kilometers thick and composed mainly of ***basalt***. ***Oceanic crust*** also has a density of about 3.0 grams per cubic centimeter. The continent's are composed of mainly ***granitic*** rock whose thickness varies between 10 and 70 kilometers. The thickest portions of ***continental crust*** are found under the various mountain ranges. Continental crust is also lighter than oceanic crust having a density of about 2.7 grams per cubic centimeter. Oceanic and continental rocks also differ from each other in terms of age. Continental crust contains some very old rocks that were formed during the ***Precambrian*** between 3 and 4 billion years ago. Oceanic rocks are normally quite young deposits. ***Isotopic dating*** of the rocks found on the sea-floor indicates that they were created less that 180 million years ago (see **topic *10i*** and ***Figure 10i-1*** for more information on the age of oceanic crust).

Variations in the age, density, and chemical composition of oceanic and continental crust suggest that these lithospheric deposits were created by different processes. The following discussion describes these differences.

Continental Crust

All of the Earth's continents have a core foundation that is made of mixtures of very old ***granite***, ***gneiss***, ***schist***, ***sedimentary***, and ***volcanic*** rocks. This core foundation is often referred to as a ***shield*** or ***basement rock***. Rocks found in the shields were formed during the ***Precambrian*** and are some of the oldest rocks found on the Earth. In Canada, some of the ***metamorphic*** rocks have been dated to an age of 3.96 billion years. Geologists believe that the major continental cores were formed by the early solidification of the lighter components of magma between 3.9 and 3.8 billion years ago. The continental shields are generally covered by younger ***sedimentary deposits***. These sedimentary rocks constitute the interior ***platforms*** of the continents. The oldest platform rocks were laid down in shallow seas about 600 million years ago. In central North America, the platform sedimentary deposits are between 1000 to 2000 meters thick. Together the shield and platform form what geologists call a ***craton***. Most of the Earth's continental cratons have been ***tectonically*** stable and have experienced little deformation for hundreds of millions of years.

Around the edge of the continental cratons are the ***continental margins***. The continental margins are primarily composed of sedimentary rocks. These sedimentary rocks were originally laid down in the oceans. ***Tectonic*** collisions and plate ***subduction*** caused the accretion of these deposits at the edges of the continental cratons (**Figure 10j-1**). In some cases, this accretion is modified by the processes of tectonic compression, ***folding***, and ***faulting*** to produce mountain ranges.

Igneous Activity and the Continents

Materials are also added to continental crust through ***intrusive*** and ***extrusive igneous*** activity. ***Plumes*** of ***magma*** from the Earth's ***asthenosphere*** are generated from the friction produced at the contact zone where oceanic crust slides past continental crust (**Figure 10j-1**). These plumes then rise upward into continental crust to form ***granitic plutons*** or a variety of ***volcanic*** features on the Earth's surface. A ***pluton*** can be defined as any ***igneous intrusion*** of rock that forms a kilometer or more below the Earth's surface. The **diagram** (**Figure 10j-2**) below illustrates some of the features associated with igneous intrusions or plutons. Some of the major features include:

Dyke: thin vertical veins of igneous rock that form in the fractures found within the crust. Because these intrusive features cool quickly their rocks are dominated with fine mineral grains.

Sill: horizontal planes of solidified magma that run parallel to the grain of the original rock deposit.

Batholith: large plutonic masses of intrusive rock with more than 100 square kilometers of surface area.

Volcanic Pipe: if a dyke reaches the surface of the Earth it is then called a ***volcanic* pipe**. Igneous deposits produced by this feature are ***extrusive*** in nature.

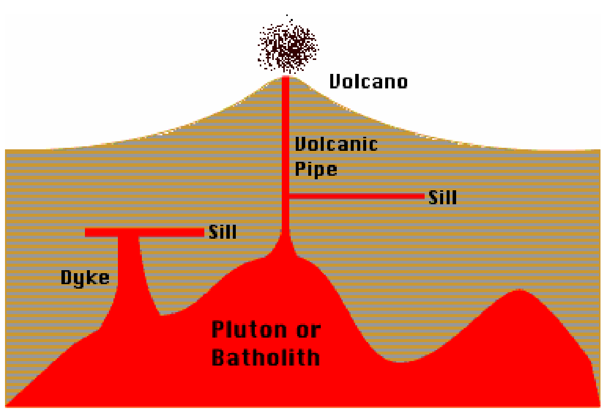
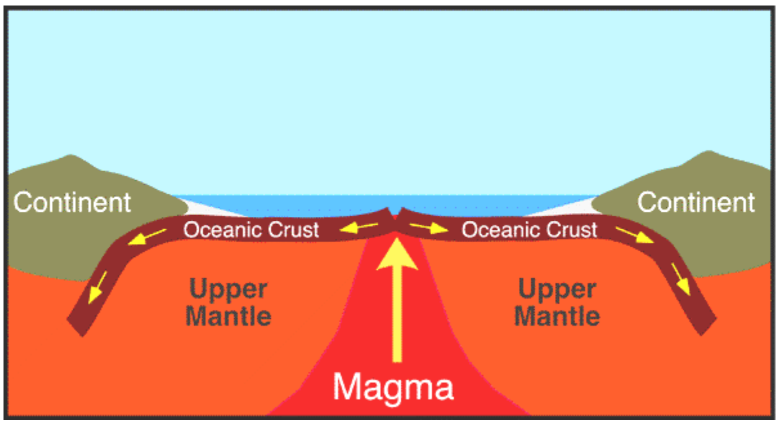


Figure 10j-2: Common plutonic features.

Plumes that are able to reach the Earth's surface produce ***volcanoes***. Most of the continental volcanoes found on our planet are located along the edge of the continents where oceanic crust is being actively subducted. In North America, the zone of active volcanoes is located on the west coast where the subduction of the Pacific plate occurs. Volcanoes add mass to the continents when magma produces ***lava flows***, ***tephra***, and ***volcanic ash***.

Oceanic Crust

Unlike continental crust, oceanic crust is actively being created at the various ***mid-oceanic ridges***. At the mid-ocean ridges, magma erupts onto the ocean floor in centrally located ***rift zones*** (**Figure 10j-3**). The newly added rock then horizontally pushes previously created ocean crust away from the rift in a conveyor belt fashion. Because of this process, we find that the age of oceanic crust increases as we move away from the rift zone (see ***Figure 10i-1***). When the oceanic crust encounters a slab of continental crust it becomes ***subducted*** because of its greater density. This process causes the oceanic crust to return to the ***mantle*** were it is re-melted into ***magma***. The process also causes the movement of continental crust across the surface of the Earth.

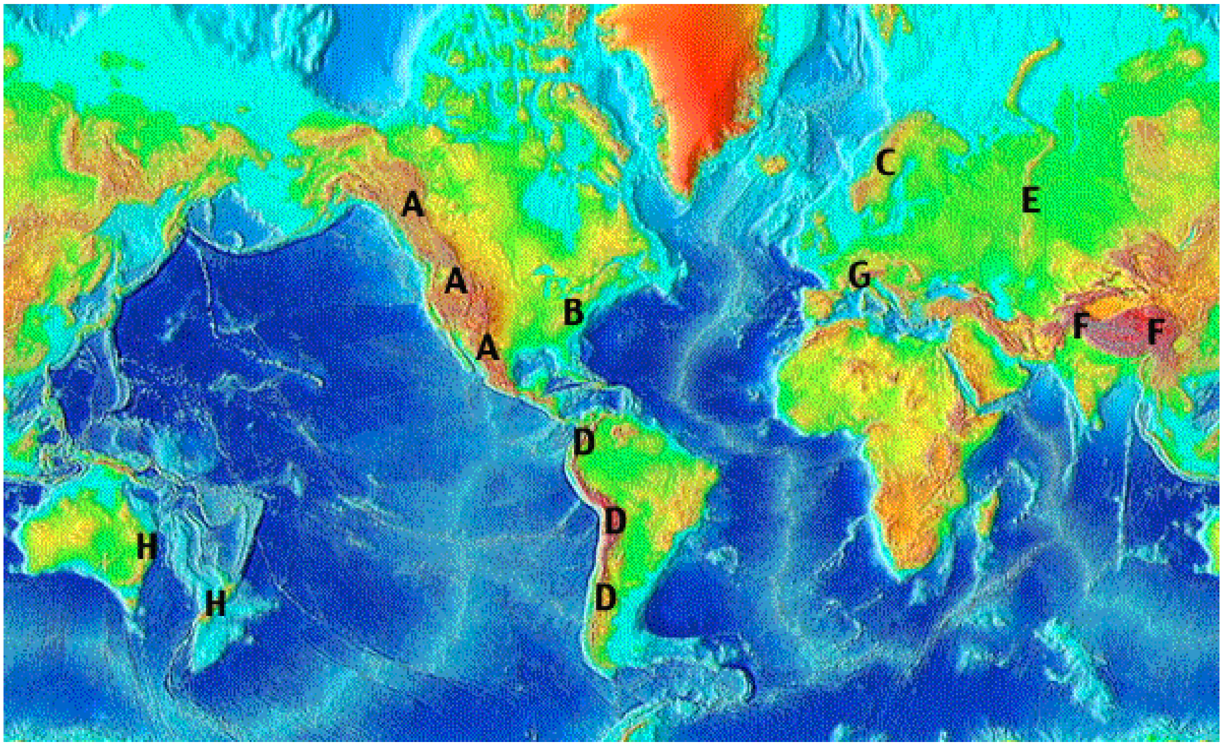


{PRIVATE**}Figure 10j-3:** Formation of oceanic crust at ***rift zones*** located in the ***mid-oceanic ridges***. Beneath the rift zone upwellings of ***magma*** occur in the ***mantle***. These upwellings produce ***fissures*** and ***volcanoes*** on the ocean floor surface. The added rock, produced from the solidification of magma, pushes previously formed oceanic crust horizontally away from the rift zone like a conveyor belt. Ocean crust is returned to the mantle through ***subduction***. This can occur when ocean crust meets continental crust or other ocean plates.

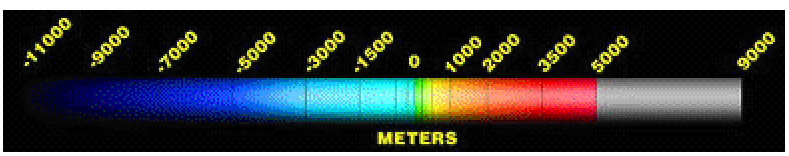
(k) Mountain Building

A **mountain** can be defined as an area of land that rises abruptly from the surrounding region. A **mountain range** is a succession of many closely spaced mountains covering a particular region of the Earth. **Mountain belts** consist of several mountain ranges that run roughly parallel to each other. The North American Cordillera, the Himalayas, the Alps, and the Appalachians are all examples of mountain belts that are composed of numerous mountain ranges.

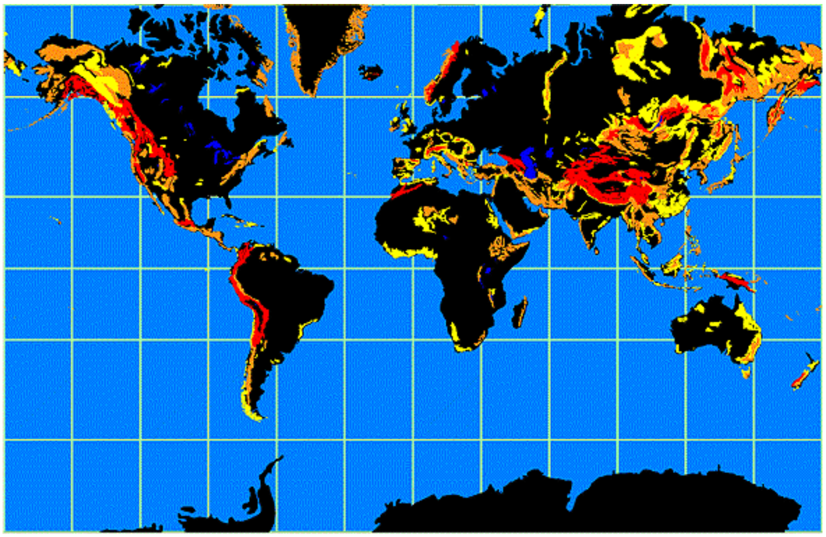
Some mountains are ***volcanic*** in origin forming where rising magma breaks through the Earth's surface. Volcanic mountains tend to have sporadic distributions within a mountain range (Mount St. Helens, Rainier, and Baker) or can occur alone because of a localized ***hot spot*** (Hawaiian Islands). Most mountains were created from tectonic forces that elevate, ***fold***, and ***fault*** rock materials. Tectonic mountains can occur as a single range (the Urals) or as a belt of several mountain ranges (North American Cordillera). **Figures 10k-1** and **10k-2** show the location of some of the major mountain systems found on the Earth's surface. These major mountain systems include the North American Cordillera, Andes, Alps, Urals, Appalachians, Himalaya, Caledonian Belt, and the Tasman Belt.



{PRIVATE**}Figure 10k-1:** This image shows the topography of both land and ocean surfaces. Elevation is indicated by color. The legend below shows the relationship between color and elevation. The Earth's major mountain systems are generally colored orange to red to grey. Some of the major mountain belts on the Earth are the North American Cordillera (**A**), Appalachians (**B**), Caledonian Belt (**C**), Andes (**D**), Urals (**E**), Himalaya (**F**), Alps (**G**), and the Tasman Belt (**H**).



The Earth's mountain ranges have various ages of formation. Parts of the Himalayas are relatively quite young. Mountain building in this region of the world began about 45 million years ago when the continental plates of India and Eurasia converged on each other. The Himalaya mountains are still actively being uplifted. The Appalachian belt is quite old. Mountain building in this region of the world started about 450 million years ago. ***Orogeny*** stopped in the Appalachians about 250 million years ago. The long passage of time without active uplift has allowed ***weathering*** and ***erosion*** to remove large amounts of bedrock from the Appalachians. These processes have also significantly lowered and rounded the peaks of the various mountains found in this belt. Mountain building episodes in the North American Cordillera have been occurring over a very long period of time and still continue today. Some ***sedimentary rocks*** in the Rocky Mountain range (located on the eastern edge of the North American Cordillera) date to over a billion years old.

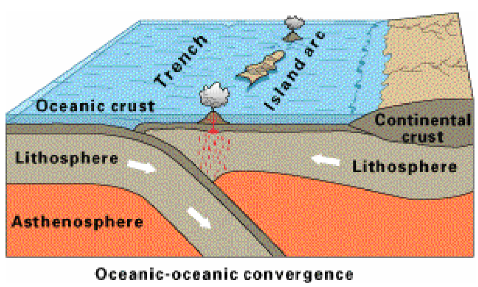


{PRIVATE**}Figure 10k-2:** The following illustration classifies the Earth's mountainous areas by elevation. The three elevation categories in this classification are: high (red), middle (orange), and low (yellow).

Evolution of Mountains

Geologists have developed a general model to explain how most mountain ranges form. This model suggests that mountain building involves three stages: (1) accumulation of sediments, (2) an orogenic period of rock deformation and crustal uplift, and (3) a period of crustal uplift caused by ***isostatic rebound*** and block-faulting. The later two stages of this model involve tectonic convergence of crustal plates which provides the compressional and tensional stresses that produce rock deformation, uplift, and faulting.

Mountain belts normally contain numerous layers of ***sedimentary*** and volcanic ***igneous rocks.*** These accumulations can be several kilometers in thickness. Most of these accumulations were originally ***deposited*** in a marine environment. The beds of the sedimentary rocks are composed of particles that came from nearby terrestrial landmasses. These particles were released from rocks by ***weathering*** and then transported by ***erosional*** forces to the edge of the terrestrial ***continental crust*** (see **topic *10p***). Beyond the edge of the continents, these sediments are ***lithified*** to form ***shales***, ***limestones***, and ***sandstones*** that make up the ***continental shelves***, ***slopes***, and ***rise***. Accumulations of volcanic rock develop along convergent boundaries where subduction is causing ***magma plumes*** to form ***plutons*** and ***volcanoes***. The volcanoes are usually spatially organized in a line, called an ***island arc***, that runs at right angles to the direction of crustal movement.

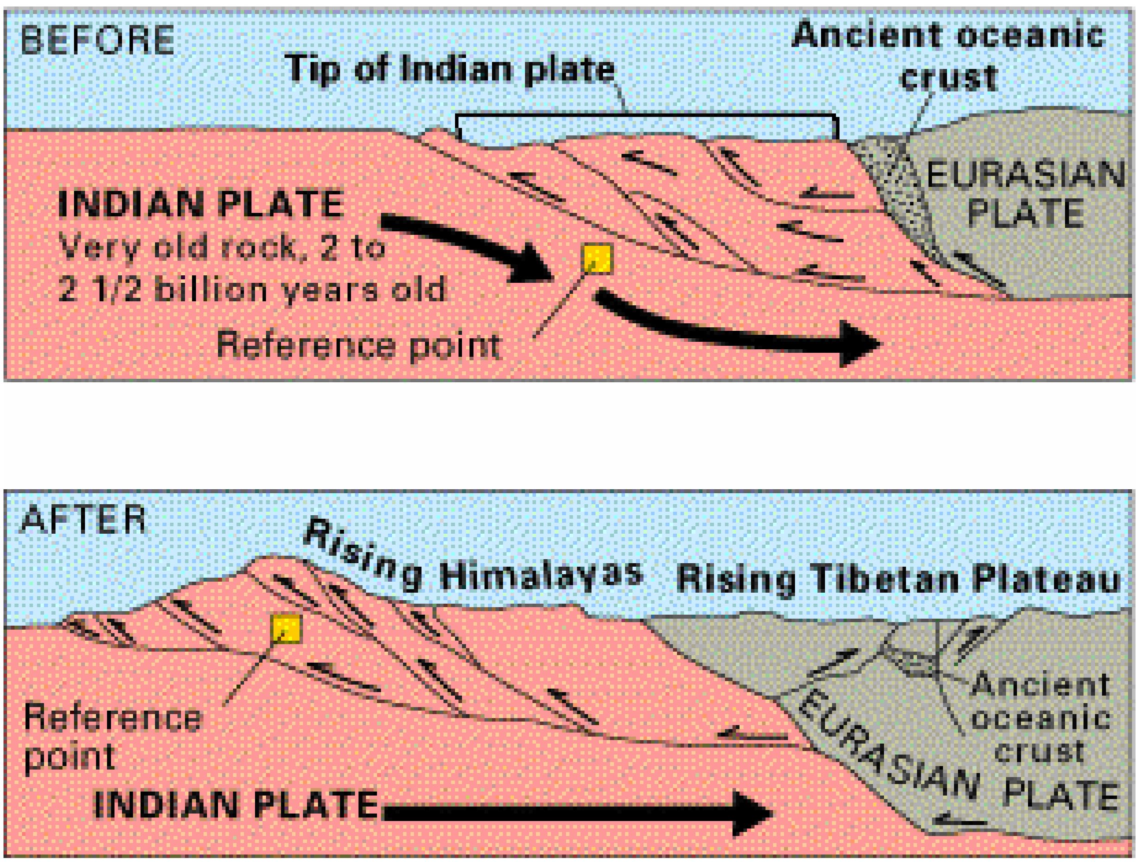


{PRIVATE**}Figure 10k-3:** Volcanic rocks found in mountains often originate from magma plumes that have migrated up through oceanic crust. The subduction of one oceanic plate under another creates friction that melts rock into magma. This magma then migrates upward through the crust forming ***plutons*** and ***volcanoes***. Volcanic rocks can also be converted into sedimentary sandstones. Weathering and erosion can remove material from terrestrial volcanic deposits to marine depositional environments. Overtime these sediments can then become lithified. (**Source:** U.S. Geological Survey).

In the **orogenic stage** of mountain building, the accumulated sediments become deformed by compressional forces from the collision of ***tectonic plates***. This tectonic convergence can be of three types: **arc-continent**, **ocean-continent** or **continent-continent**. In an **ocean-continent** convergence, the collision of ocean and continental plates causes the ***accretion*** of marine sedimentary deposits to the edge of the continent. **Arc-continent** convergence occurs when an ***island arc*** collides with the edge of a continental plate. In this convergence, the ocean plate area between the arc and the continent is subducted into the ***asthenosphere*** and the volcanic rocks and sediments associated with the island arc become ***accreted*** to the margin of the continent over time. This type of collision may have been responsible for the creation of the Sierra Nevada mountains in California during the Mesozoic Era. The final type of convergence occurs when an ocean basin closes and two continental plates collide. **Continent-continent** convergence mountain building is responsible for the formation of the Himalayas, Ural, and Appalachian mountain systems.

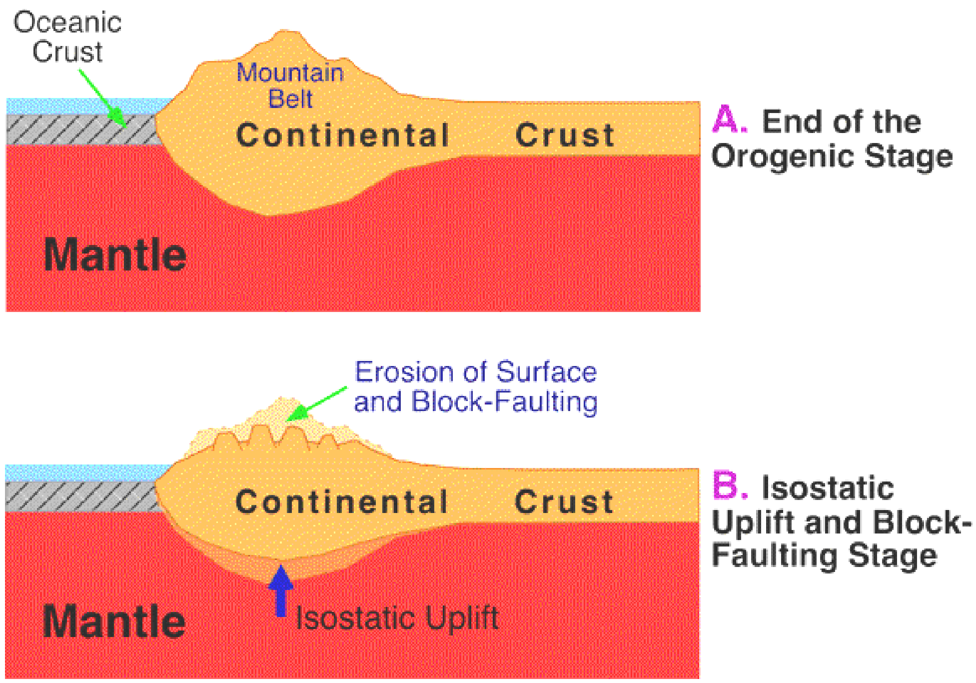
In all three types of tectonic convergence, layered rocks that were once located in the ocean basin are squeezed into a smaller and smaller area. This compression causes the once flat sedimentary beds to be ***folded*** and uplifted. When the compressional forces become greater than the rocks ability to deform, faulting occurs. Compressional forces typically result in ***reverse*** and ***overthrust faulting***. Another consequence of the orogenic stage is ***regional metamorphism*** and the incursion of ***magma plumes***, ***plutons***, and ***volcanoes*** into the growing mountain range.

Figure 10k-4 illustrates how the collision Eurasian and Indian plates created the Himalayas. In this ***orogeny***, compressional forces squished sedimentary deposits that existed between the converging continental plates and rocks at the margin of the Eurasian and Indian plates upward in elevation. These forces also created a number of overthrust faults.



{PRIVATE}**Figure 10k-4:** Formation of the Himalaya Mountains. Compressional forces due to the collision of the Eurasian and Indian continental plates caused ocean sediments and continental rocks to be pushed upward in elevation. (**Source:** U.S. Geological Survey).

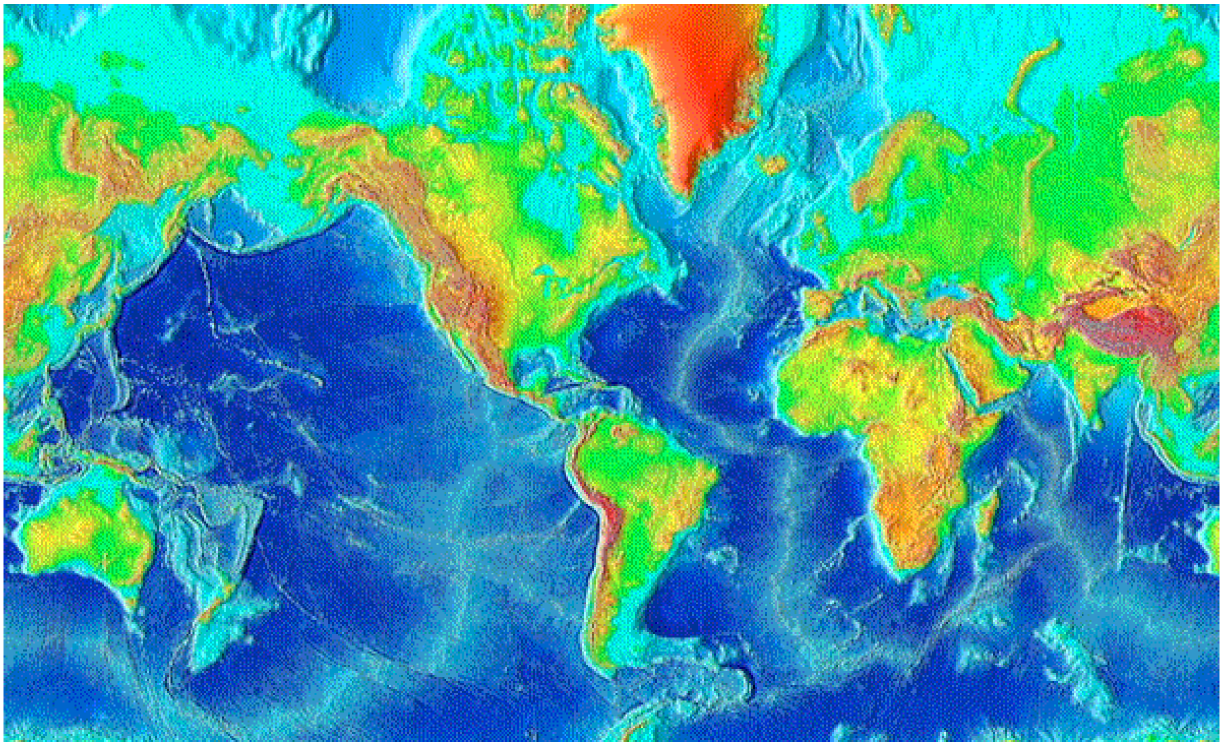
At the end of plate convergence, mountain building enters its final stage. This stage is characterized by crustal uplift because of ***isostatic rebound*** and block-faulting (**Figure 10k-5**). Isostatic rebound involves the vertical movement of continental crust that is floating in the plastic upper ***mantle***. As erosion removes surface materials from mountains, the weight of the crust in this region becomes progressively less. With less weight, the continental crust makes an isostatic adjustment causing it to rise vertically (float higher) in the mantle. This process also causes tensional forces to exist in a horizontal direction breaking the continental crust into a number of blocks. Each block moves vertically to compensate for the tensional forces producing ***normal*** and ***graben faults*** (see **topic *10l***).



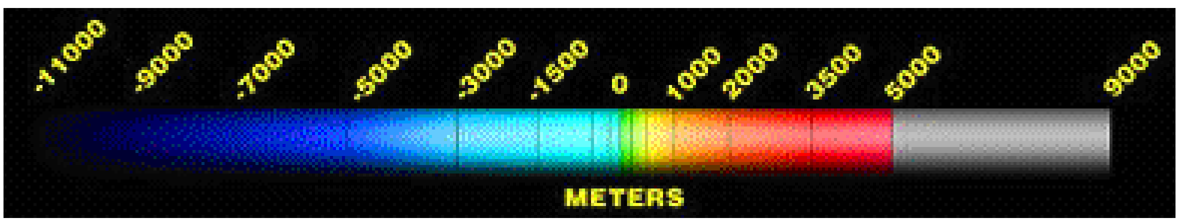
{PRIVATE**}Figure 10k-5:** After the **orogenic stage**, weathering and erosion begin removing material from the surface of the newly created mountains. The removal of rock mass makes the area of the continental crust where the mountains are less heavy and that end of the crust begins to float higher in the mantle. This ***isostatic rebound*** causes vertical uplift and the tensional forces due to the movement of the crust creates ***normal*** and ***graben faults***.

(l) Crustal Deformation Processes: Folding and Faulting

{PRIVATE**}**The topographic map illustrated in **Figure 10l-1** suggests that the Earth's surface has been deformed. This deformation is the result of forces that are strong enough to move ocean sediments to an elevation many thousands meters above sea level. In previous lectures, we have discovered that this displacement of rock can be caused by ***tectonic plate movement*** and ***subduction***, ***volcanic activity***, and ***intrusive igneous activity***.



{PRIVATE**}Figure 10l-1:** Topographic relief of the Earth's terrestrial surface and ocean basins. Ocean trenches and the ocean floor have the lowest elevations on the image and are colored dark blue. Elevation is indicated by color. The legend below shows the relationship between color and elevation. (**Source:** *National Geophysical Data Center*, National Oceanic and Atmospheric Administration).



Deformation of rock involves changes in the shape and/or volume of these substances. Changes in shape and volume occur when stress and strain causes rock to buckle and fracture or crumple into folds. A ***fold*** can be defined as a bend in rock that is the response to compressional forces. Folds are most visible in rocks that contain layering. For plastic deformation of rock to occur a number of conditions must be met, including:

The rock material must have the ability to deform under pressure and heat.

The higher the temperature of the rock the more plastic it becomes.

Pressure must not exceed the internal strength of the rock. If it does, fracturing occurs.

Deformation must be applied slowly.

A number of different **folds** have been recognized and classified by geologists. The simplest type of fold is called a ***monocline*** (**Figure 10i-2**). This fold involves a slight bend in otherwise parallel layers of rock.

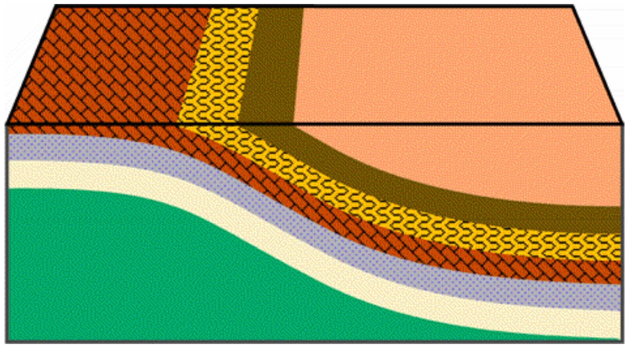
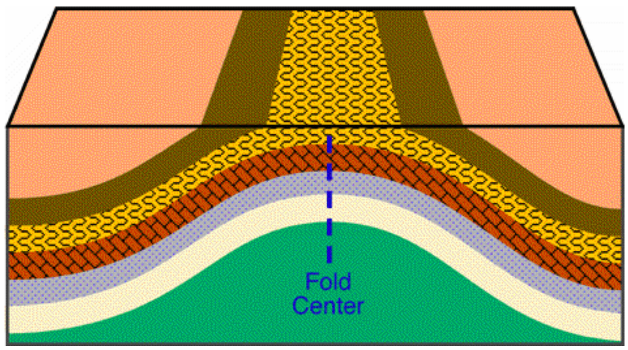


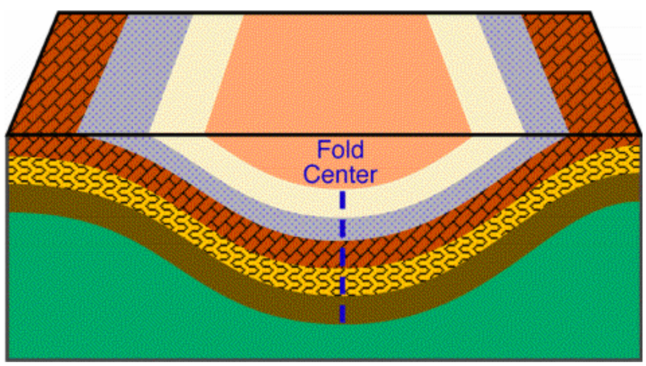
Figure 10l-2: Monocline fold.

An ***anticline*** is a convex up fold in rock that resembles an arch like structure with the rock beds (or limbs) ***dipping*** way from the center of the structure (**Figure 10l-3**).



{PRIVATE**}Figure 10l-3:** Anticline fold. Note how the rock layers dip away from the center of the fold are roughly symmetrical.

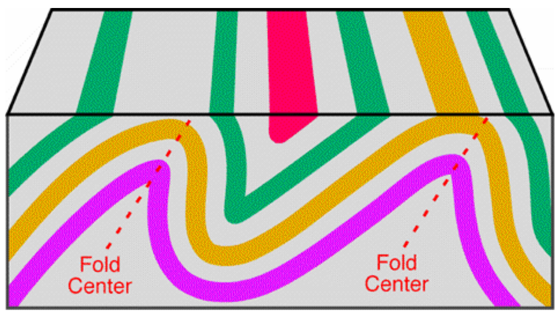
A ***syncline*** is a fold where the rock layers are warped downward (**Figure 10l-4** and **10l-5**). Both anticlines and synclines are the result of compressional stress.



{PRIVATE**}Figure 10l-4:** Syncline fold. Note how the rock layers dip toward the center of the fold and are roughly symmetrical.

{PRIVATE**}Figure 10l-5:** Synclinal folds in bedrock, near Saint-Godard-de-Lejeune, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

More complex fold types can develop in situations where lateral pressures become greater. The greater pressure results in **anticlines** and **synclines** that are inclined and asymmetrical (**Figure 10l-6**).



{PRIVATE**}Figure 10l-6:** The following illustration shows two anticline folds which are inclined. Also note how the beds on either side of the fold center are asymmetrical.

A ***recumbent fold*** develops if one limb of the fold passes the vertical (**Figure 10l-7**). Recumbent folds are commonly found in the core of mountain ranges and indicate that compression and/or shear forces were stronger in one direction. Extreme stress and pressure can sometimes cause the rocks to shear along a plane of weakness creating a ***fault***. We call the combination of a **fault** and a **fold** in a rock an ***overthrust fault***.

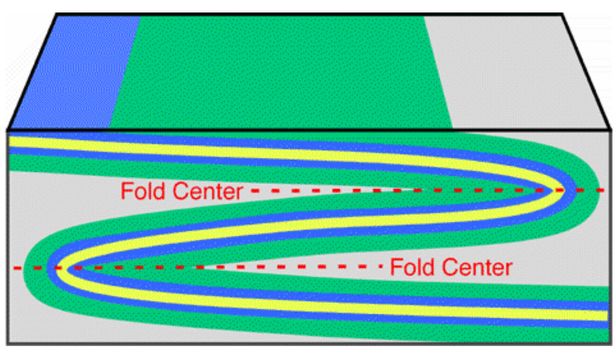
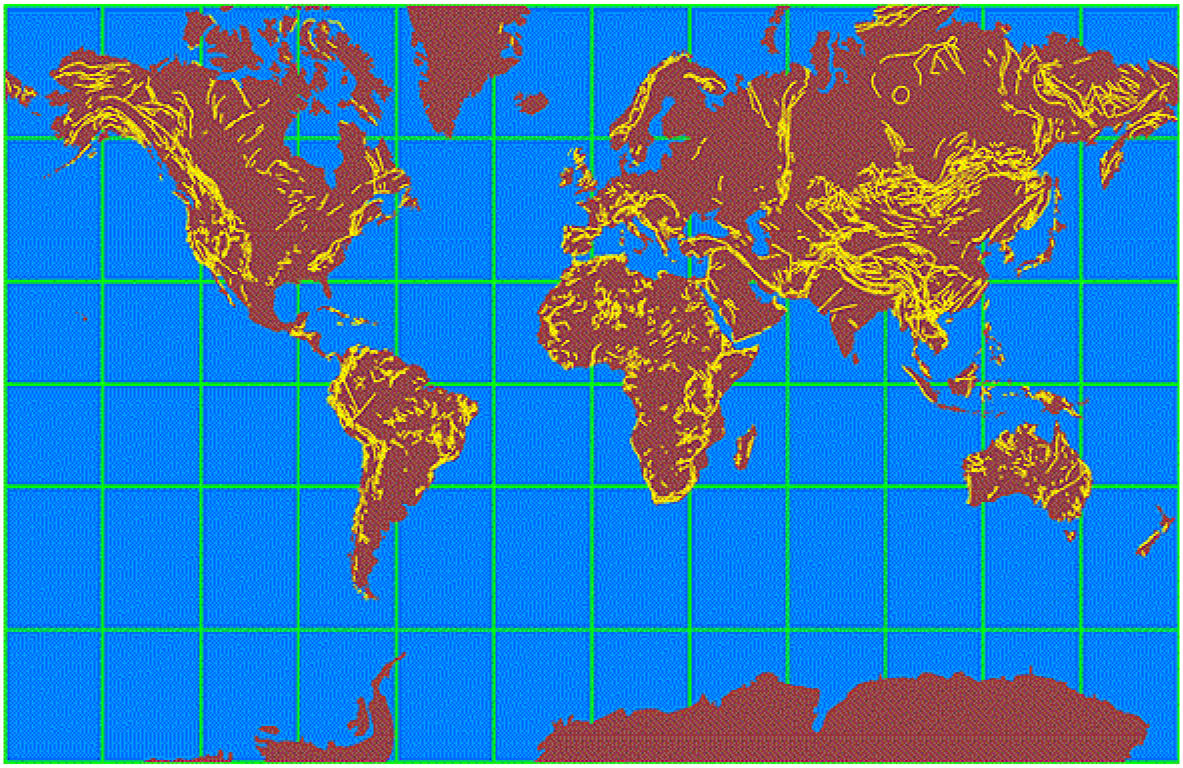


Figure 10l-7: Recumbent fold.

Faults form in rocks when the stresses overcome the internal strength of the rock resulting in a fracture. A fault can be defined as the displacement of once connected blocks of rock along a **fault plane**. This can occur in any direction with the blocks moving away from each other. Faults occur from both tensional and compressional forces. **Figure 10l-8** shows the location of some of the major faults located on the Earth.



{PRIVATE**}Figure 10l-8:** Location of some of the major faults on the Earth. Note that many of these faults are in mountainous regions (see ***section 10k***).

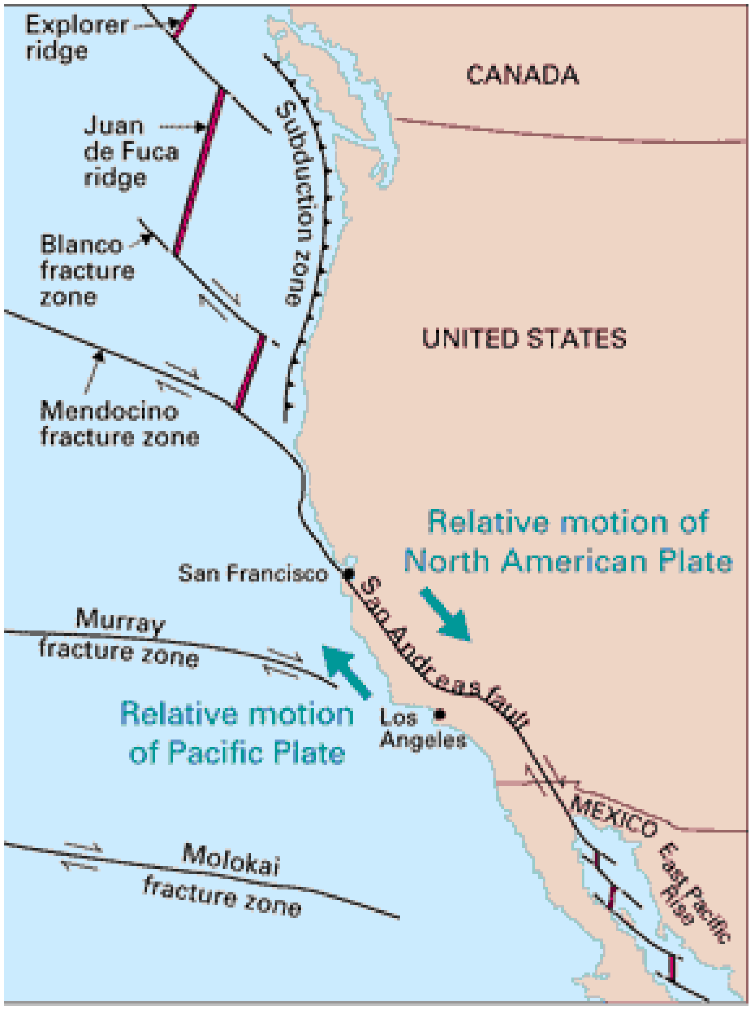
There are several different kinds of faults. These faults are named according to the type of stress that acts on the rock and by the nature of the movement of the rock blocks either side of the **fault plane**. ***Normal faults*** occur when tensional forces act in opposite directions and cause one slab of the rock to be displaced up and the other slab down (**Figure 10l-9**).

Reverse faults develop when compressional forces exist (**Figure 10l-10**). Compression causes one block to be pushed up and over the other block.

A ***graben fault*** is produced when tensional stresses result in the ***subsidence*** of a block of rock. On a large scale these features are known as ***Rift Valleys*** (**Figure 10l-11**).

A ***horst fault*** is the development of two **reverse faults** causing a block of rock to be pushed up (**Figure 10l-12**).

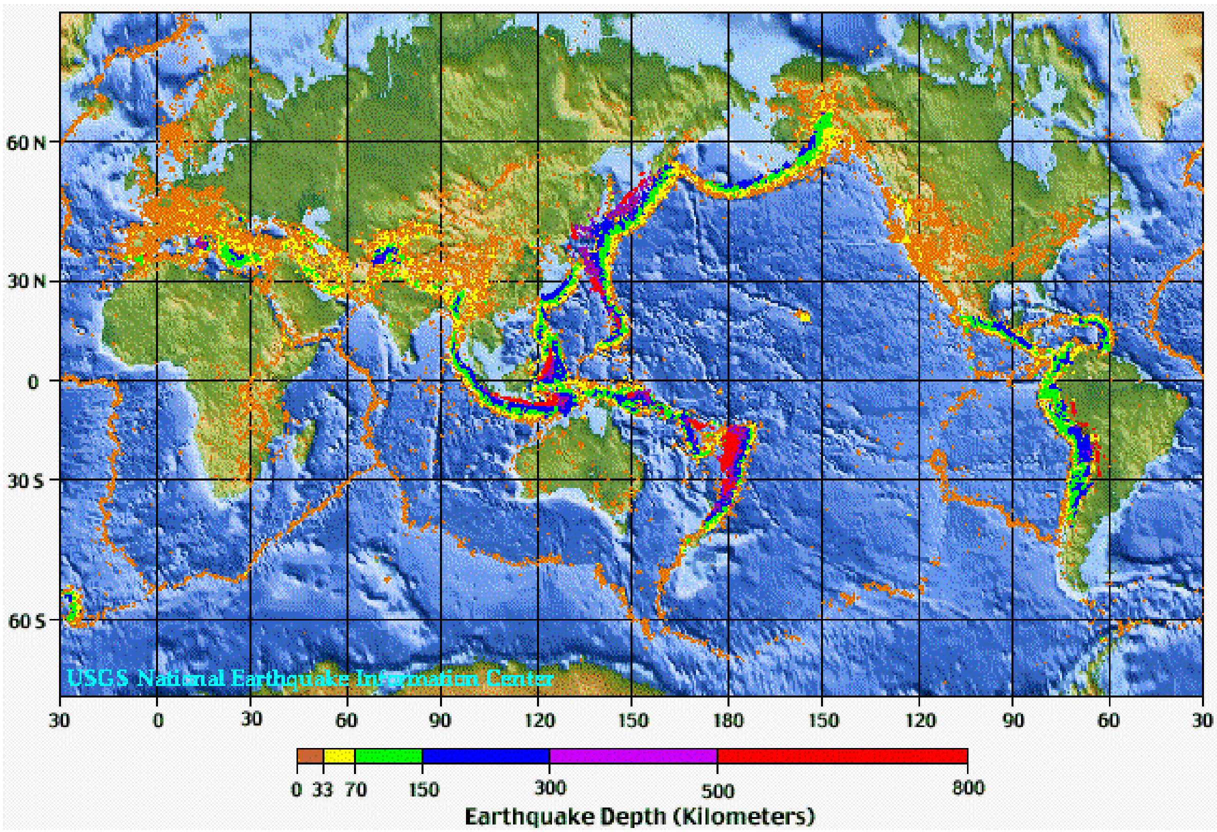
The final major type of fault is the ***strike-slip*** or ***transform fault***. These faults are vertical in nature and are produced where the stresses are exerted parallel to each other (**Figure 10l-13**). A well-known example of this type of fault is the **San Andreas fault** in California.



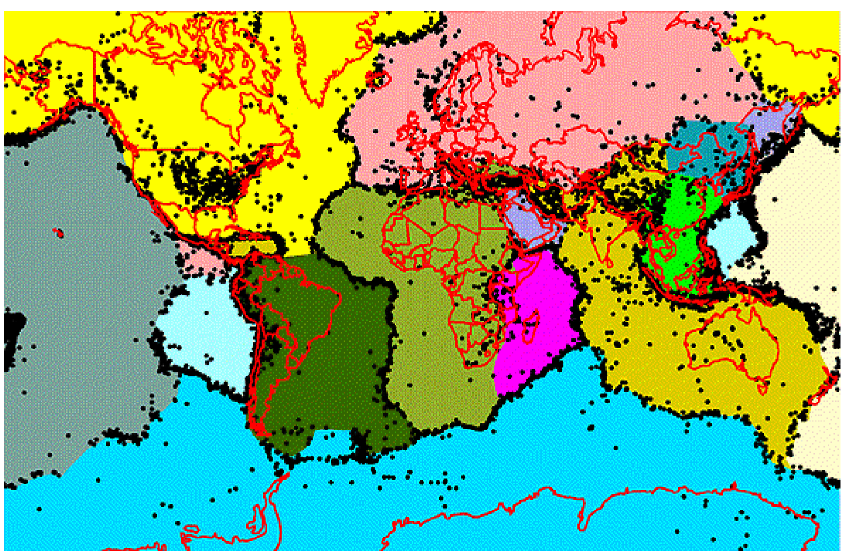
{PRIVATE**}Figure 10l-13:** Transcurrent fault zones on and off the West coast of North America. (**Source:** U.S. Geological Survey).

(m) Earthquakes

An ***earthquake*** is a sudden vibration or trembling in the Earth. More than 150,000 tremors strong enough to be felt by humans occur each year worldwide. Earthquake motion is caused by the quick release of stored ***potential energy*** into the ***kinetic energy*** of motion. Most earthquakes are produced along ***faults***, ***tectonic plate*** boundary zones, or along the ***mid-oceanic ridges*** (**Figures 10m-1** and **10m-2**). At these areas, large masses of rock that are moving past each other can become locked due to ***friction***. Friction is overcome when the accumulating stress has enough force to cause a sudden slippage of the rock masses. The magnitude of the shock wave released into the surrounding rocks is controlled by the quantity of stress built up because of friction, the distance the rock moved when the slippage occurred, and ability of the rock to transmit the energy contained in the ***seismic waves***. The San Francisco earthquake of 1906 involved a 6 meter horizontal displacement of bedrock. Sometime after the main shock wave, ***aftershocks*** can occur because of the continued release of frictional stress. Most aftershocks are smaller than the main earthquake, but they can still cause considerable damage to already weakened natural and human constructed features.



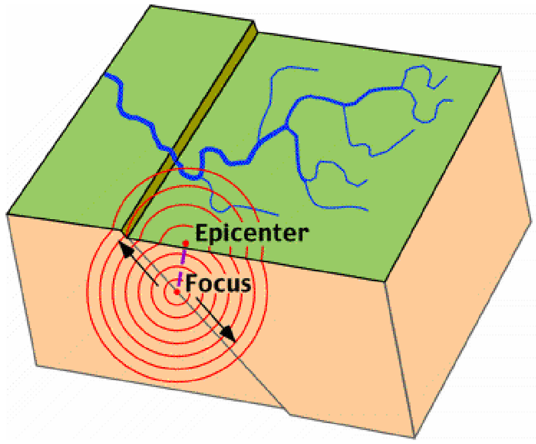
{PRIVATE**}Figure 10m-1:** Distribution of earthquake ***epicenters*** from 1975 to 1995. Depth of the ***earthquake focus*** is indicated by color. Deep earthquakes occur in areas where oceanic crust is being actively subducted. About 90 % of all earthquakes occur at a depth between 0 and 100 kilometers. (**Source:** U.S. Geologic Survey, *National Earthquake Information Center*).



{PRIVATE**}Figure 10m-2:** Distribution of earthquakes with a magnitude less than 5.0 relative to the various tectonic plates found on the Earth's surface. Each tectonic plate has been given a unique color. This illustration indicates that the majority of small earthquakes occur along plate boundaries.

Earthquake Waves

Earthquakes are a form of **wave energy** that is transferred through bedrock. Motion is transmitted from the point of sudden energy release, the ***earthquake focus***, as spherical ***seismic waves*** that travel in all directions outward (**Figure 10m-3**). The point on the Earth's surface directly above the focus is termed the ***epicenter***.



{PRIVATE**}Figure 10m-3:** Movement of body waves away from the focus of the earthquake. The epicenter is the location on the surface directly above the earthquake's focus.

Two different types of seismic waves have been described by geologists: ***body waves*** and ***surface waves***. Body waves are seismic waves that travel through the lithosphere. Two kinds of body waves exist: ***P-waves*** and ***S-waves***. Both of these waves produce a sharp jolt or shaking. P-waves or **primary waves** are formed by the alternate expansion and contraction of bedrock and cause the volume of the material they travel through to change (**Figure 10m-4**). They travel at a speed of about 5 to 7 kilometers per second through the ***lithosphere*** and about 8 kilometers per second in the ***asthenosphere***. The speed of sound is about 0.30 kilometers per second. P-waves also have the ability to travel through solid, liquid, and gaseous materials. When some P-waves move from the ground to the lower atmosphere, the sound wave that is produced can sometimes be heard by humans and animals.

S-waves or **secondary waves** are a second type of body wave. These waves are slower than P-waves and can only move through solid materials. S-waves are produced by ***shear stresses*** and move the materials they pass through in a perpendicular (up and down or side to side) direction.

Surface waves travel at or near the Earth's surface. These waves produce a rolling or swaying motion causing the Earth's surface to behave like waves on the ocean. The velocity of these waves is slower than body waves. Despite their slow speed, these waves are particularly destructive to human construction because they cause considerable ground movement.

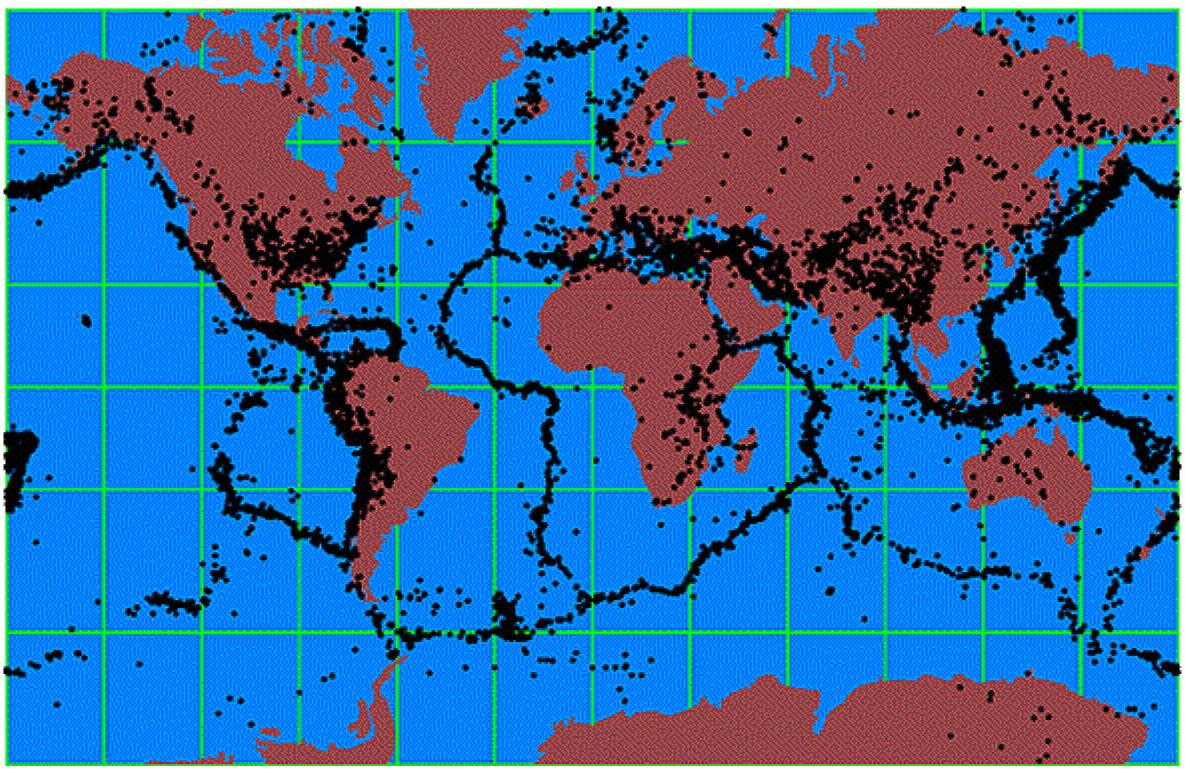
Earthquake Measurement

The strength of an earthquake can be measured by a device called a ***seismograph***. When an earthquake occurs this device converts the wave energy into a standard unit of measurement like the ***Richter scale***. In the Richter scale, units of measurement are referred to as ***magnitudes***. The Richter scale is ***logarithmic***. Thus, each unit increase in magnitude represents 10 times more energy released. **Table 10m-1** describes the relationship between Richter scale magnitude and energy released.

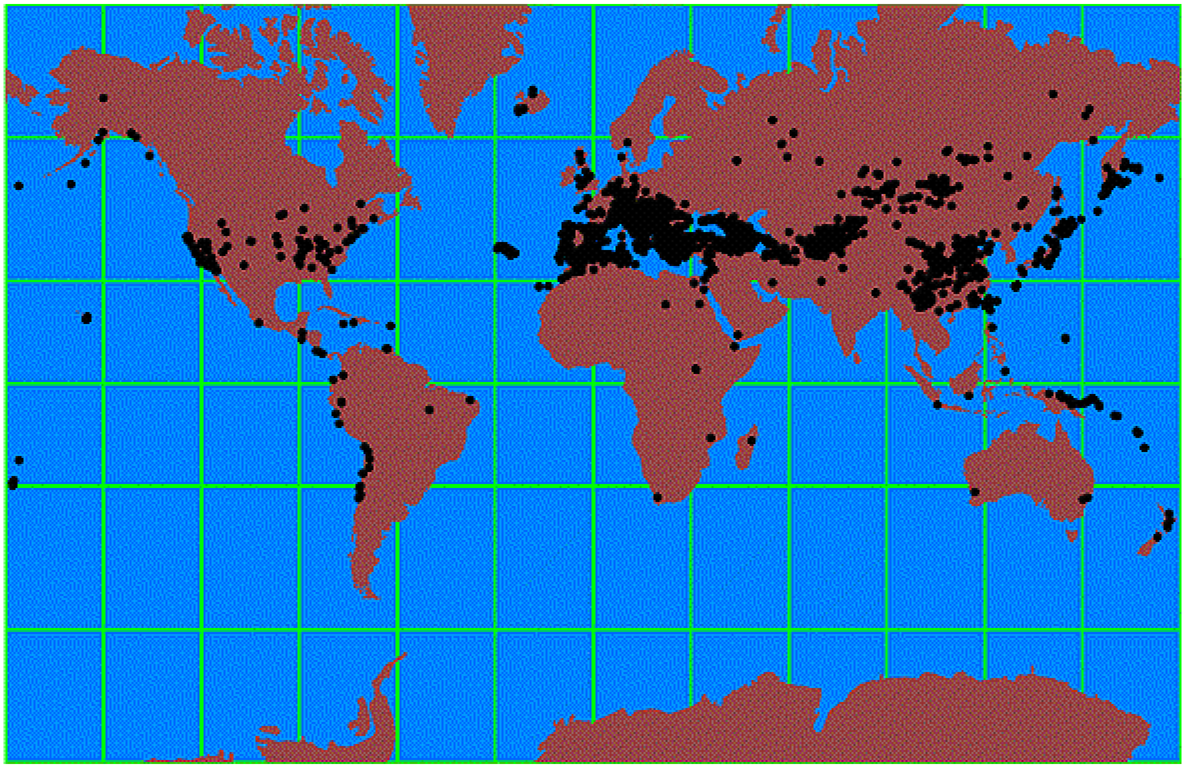
Table 10m-1: Relationship between Richter Scale magnitude and energy released.

{PRIVATE}**Magnitude in Richter Scale**	**Energy Released in Joules**	**Comment**
2.0	6.3×10^{7}	Smallest earthquake detectable by people.
5.0	2.0×10^{12}	Energy released by the Hiroshima atomic bomb.
6.0 - 6.9	6.3×10^{13} to 1.4×10^{15}	About 120 shallow earthquakes of this magnitude occur each year.
6.7	7.1×10^{14}	Northridge, California earthquake 1994.
7.0	2.0×10^{15}	Major earthquake.
7.4	7.9×10^{15}	Turkey earthquake August 17, 1999. More than 12,000 people killed.
7.6	1.6×10^{16}	Deadliest earthquake this century. Tangshan, China, 1976. About 250,000 people died.
8.3	1.8×10^{17}	San Francisco earthquake of 1906.
8.6	5.0×10^{17}	Most powerful earthquake recorded in the last 100 years. Southern Chile 1960. Claimed 5,700 lives.

Figures 10m-5 and **10m-6** describe the spatial distribution of small and large earthquakes respectively. These maps indicate that large earthquakes have distributions that are quite different from small events. Many large earthquakes occur some distance away from a plate boundary. Some geologists believe that these powerful earthquakes may be occurring along ancient faults that are buried deep in the continental crust. Recent seismic studies in the central United States have discovered one such fault located thousands of meters below the lower Mississippi Valley. Some large earthquakes occur at particular locations along the plate boundaries. Scientists believe that these areas represent zones along adjacent plates that have greater frictional resistance and stress.



{PRIVATE}**Figure 10m-5:** Distribution of earthquakes with a magnitude less than 5 on the Richter Scale.



{PRIVATE**}Figure 10m-6:** Distribution of earthquakes with a magnitude greater than 7 on the Richter Scale.

Earthquake Damage and Destruction

Earthquakes are a considerable hazard to humans. Earthquakes can cause destruction by structurally damaging buildings and dwellings, fires, ***tsunamis***, and ***mass wasting*** (see **Figures 10m-7** to **10m-11**). Earthquakes can also take human lives. The amount of damage and loss of life depends on a number of factors. Some of the more important factors are:

Time of day. Higher losses of life tend to occur on weekdays between the hours of 9:00 AM to 4:00 PM. During this time interval many people are in large buildings because of work or school. Large structures are often less safe than smaller homes in an earthquake.

Magnitude of the earthquake and duration of the event.

Distance form the earthquake's focus. The strength of the shock waves diminish with distance from the focus.

Geology of the area effected and soil type. Some rock types transmit seismic wave energy more readily. Buildings on solid bedrock tend to receive less damage. Unconsolidated rock and sediments have a tendency to increase the amplitude and duration of the seismic waves increasing the potential for damage. Some soil types when saturated become liquefied (**Figure 10m-7**).

Type of building construction. Some building materials and designs are more susceptible to earthquake damage (**Figure 10m-8**).

Population density. More people often means greater chance of injury and death.

The greatest loss of life because of an earthquake this century occurred in Tangshan, China in 1976 when an estimated 250,000 people died. In 1556, a large earthquake in the Shanxi Province of China was estimated to have caused the death of about 1,000,000 people.

{PRIVATE**}Figure 10m-7:** Earthquake of June 16, 1964 in Niigata, Japan had a magnitude of 7.4. ***Liquefaction*** of some soils in the area caused large apartment buildings to tip over on their sides. (**Source:** Image provided by the *National Geophysical Data Center*, NOAA).

{PRIVATE**}Figure 10m-8:** A view of a parking lot on the campus of California State University. Columns of reinforced concrete failed after the 1994 Northridge earthquake and its aftershocks. (**Source:** Photography by M. Celebi, US Geological Survey. Image provided by the *National Geophysical Data Center*, NOAA).

A common problem associated with earthquakes in urban areas is fire (**Figure 10m-9**). Shaking and ground displacement often causes the severing of electrical and gas lines leading to the development of many localized fires. Response to this problem is usually not effective because shock waves also rupture pipes carrying water. In the San Francisco earthquake of 1906, almost 90 % of the damage to buildings was caused by fire.

{PRIVATE**}Figure 10m-9:** The following image looks at downtown Kobi, Japan at about noon on the day of the 1995 earthquake. Many areas of downtown Kobi were on fire and there was no water pressure to put out the flames. (**Source:** Photograph by Roger Hutchison. Image provided by the *National Geophysical Data Center*, NOAA).

In mountainous regions, earthquake provoked landslides can cause many deaths and severe damage to built structures (**Figure 10m-10**). The town of Yungay, Peru was buried by a ***debris flow*** that was triggered by an earthquake that occurred on May 31, 1970. This disaster engulfed the town in seconds with mud, rock, ice, and water and took the lives of about 20,000 people.

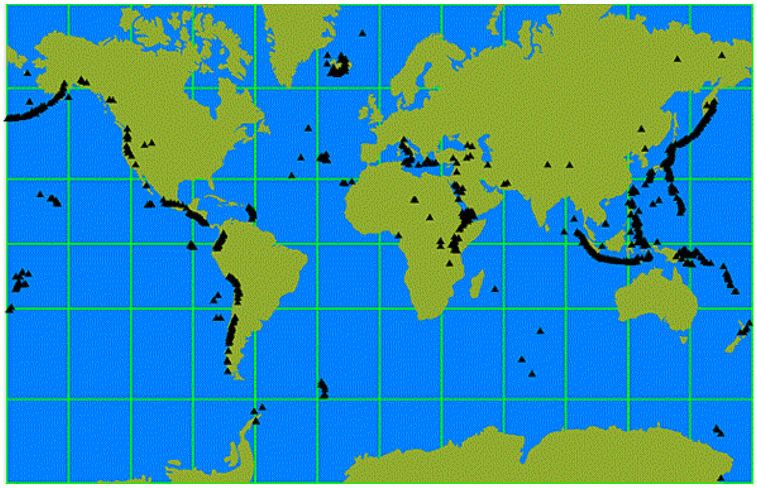
{PRIVATE**}Figure 10m-10:** The Guatemala earthquake of February 4, 1976 had a magnitude of 7.5. This earthquake killed about 23,000 people, injured 76,000, and caused just over 1 billion dollars in property damage. The earthquake also caused a number of landslides. (**Source:** Photography by US Geological Survey. Image provided by the *National Geophysical Data Center*, NOAA).

Another consequence of earthquakes is the generation of ***tsunamis*** (**Figure 10m-11**). Tsunamis, or tidal waves, form when an earthquake causes a sudden movement of the seafloor. This movement creates a wave in the water body which radiates outward in concentric shells. On the open ocean, these waves are usually no higher than one to three meters in height and travel at speed of about 750 kilometers per hour. Tsunamis become dangerous when they approach land. Frictional interaction of the waves with the ocean floor, as they near shore, causes the waves to slow down and collide into each other. This amalgamation of waves then produces a super wave that can be as tall as 65 meters in height.

{PRIVATE**}Figure 10m-11:** The earthquake of March 27, 1964, in the Gulf of Alaska generated a tsunami. This photo shows a beached fishing boat that was carried landward by the tsunami wave. (**Source:** Image provided by the *National Geophysical Data Center*, NOAA).

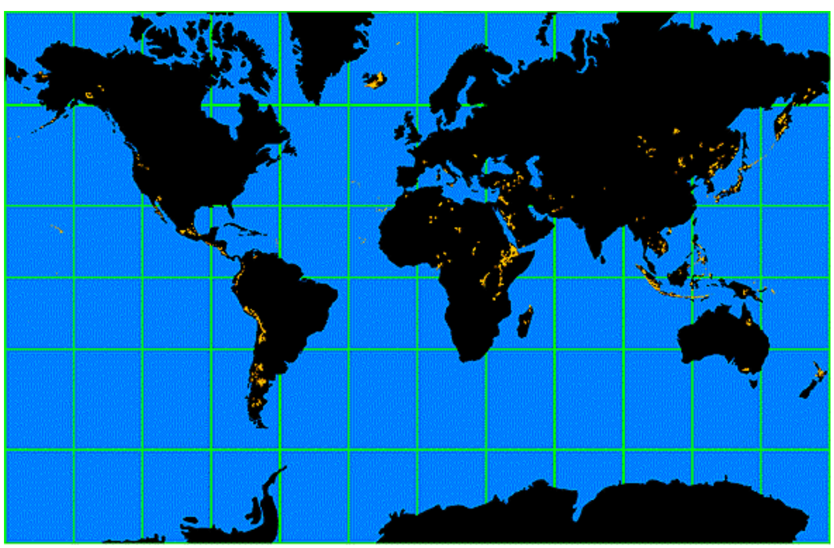
(n) Volcanism

{PRIVATE**}**A ***volcano*** is generally a conical shaped hill or mountain built by accumulations of ***lava flows***, ***tephra***, and ***volcanic ash***. About 95 % of active volcanoes occur at the plate ***subduction zones*** and at the ***mid-oceanic ridges*** (**Figure 10n-1**). The other 5 % occur in areas associated with ***lithospheric hot spots***. These hot spots have no direct relationships with areas of crustal creation or subduction zones. It is believed that hot spots are caused by ***plumes*** of rising ***magma*** that have their origin within the ***asthenosphere***.



{PRIVATE**}Figure 10n-1**: Location of the Earth's major volcanoes. Most occur along tectonic plate boundaries where plate ***subduction*** creates rising ***plumes*** of ***magma***. The volcanoes that do not occur along plate boundaries are the result of localized ***asthenosphere hot spots*** that melt through the Earth's crust. The Hawaiian Island chain of volcanoes was create by a hot spot.

Over the last 2 million years, volcanoes have been depositing lava, tephra, and ash in particular areas of the globe (**Figure 10n-2**). These areas occur at ***hot spots***, ***rift zones***, and along plate boundaries where tectonic ***subduction*** is taking place.



{PRIVATE**}Figure 10n-2**: Location of major volcanic deposits laid down during the last 2 million years.

Not all volcanoes are the same. Geologists have classified five different types of volcanoes. This classification is based on the geomorphic form, magma chemistry, and the explosiveness of the eruption.

The least explosive type of volcano is called a ***basalt plateau***. These volcanoes produce a very fluid ***basaltic magma*** with horizontal flows. The form of these volcanoes is flat to gently sloping and they can occupy an area from 100,000 to 1,000,000 square kilometers. Deposits of these volcanoes can be as thick as 1800 meters. Large basalt plateaus are found in the Columbia River Plateau, western India, northern Australia, Iceland, Brazil, Argentina, and Antarctica.

Some basaltic magmas can produce very large slightly sloping volcanoes, 6 to 12 degrees, that have gently flowing magmas called ***shield volcanoes*** (see ***link***). Shield volcanoes can be up to 9000 meters tall. The volcanoes of the ***Hawaiian Islands*** are typical of this type. Extruded materials from this type of volcano mainly consist of low viscosity basaltic lava flows (**Figure 10n-3**).

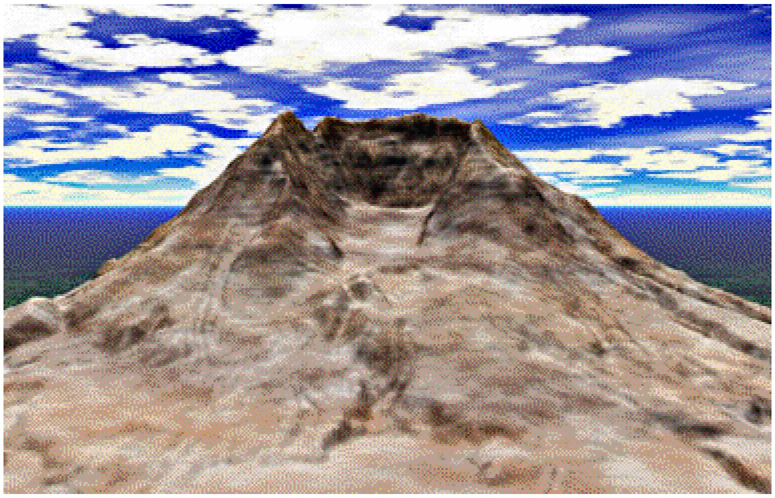
{PRIVATE**}Figure 10n-3**: Low viscosity basaltic lava flow from an active volcano on one of the Hawaiian Islands.

A ***cinder cone*** is a small volcano, between 100 and 400 meters tall, made up of exploded rock blasted out of a central vent at a high velocity (**Figure 10n-4** and see ***link***). These volcanoes develop from magma of basaltic to intermediate composition (***andesite***). They form when large amounts of gas accumulate within rising magma. Examples of cider cones include Little Lake Volcano in California and ***Paricutin Volcano*** in Mexico.

{PRIVATE**}Figure 10n-4**: Cinder cone volcano. Note how the vegetation near the mound has been burnt by lava flows.

Composite volcanoes are made from alternate layers of lava flows and exploded rock (see ***link***). Their height ranges from 100 to 3500 meters tall. The chemistry of the magma of these volcanoes is quite variable ranging from ***basalt*** to ***granite***. Magmas that are more granitic tend to be very explosive because of their relatively higher water content. Water at high temperatures and pressures is extremely volatile. Examples of composite volcanoes include Italy's ***Vesuvius***, Japan's ***Mount Fuji***, and Washington State's ***Mount Rainier*** and Mount St. Helens (see **Figures 10n-5** and **10n-6**).

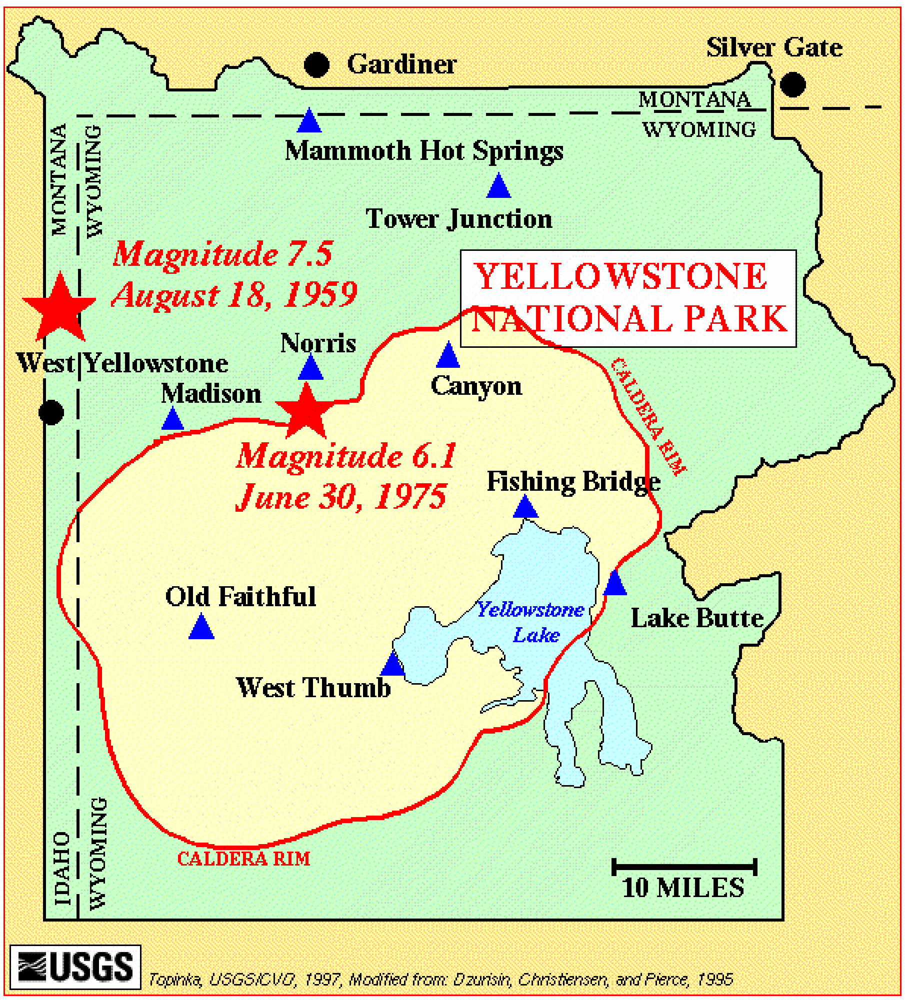
{PRIVATE**}Figure 10n-5**: Mount St. Helens eruption on May 18, 1980. (**Source:** U.S. Geological Survey, photograph by Austin Post).



{PRIVATE**}Figure 10n-6**: The above image is a post-eruption computer rendering of Mount St. Helens from a U.S.Geological Survey digital elevation model (DEM). The lateral eruption removed 2.8 cubic kilometers of rock and sediment from from the volcano and lowered its height by 400 meters. Detectable amounts of ash were spread over 50,000 square kilometers of area surrounding the volcano. The large crater created by the explosive eruption is about 600 meters deep and can be seen in the center of the image above.

The most explosive type of volcano is the ***caldera*** (see ***link***). The cataclysmic explosion of these volcanoes leaves a huge circular depression at the Earth's surface. This depression is usually less than 40 kilometers in diameter. These volcanoes form when "wet" ***granitic magma*** quickly rises to the surface of the Earth. When it gets to within a few kilometers of the surface the top of the magma cools to form a dome. Beneath this dome the gaseous water in the magma creates extreme pressures because of expansion. When the pressure becomes too great the dome and magma are sent into the Earth's atmosphere in a tremendous explosion. On the island of ***Krakatau***, a caldera type volcano exploded in 1883 ejecting 75 cubic kilometers of material in the air and left a depression in the ground some 7 kilometers in diameter.

A potentially very destructive caldera covering an area of about 2000 square kilometers exists under Yellowstone National Park in the United States (**Figure 10n-7**). Investigations have discovered that over the last 2 million years this volcano has exploded on a regular interval of about 700,000 years. The last eruption occurred 630,000 years ago and the next could take place anytime.When the Yellowstone caldera last erupted, it blasted 1,000 cubic kilometers of volcanic ash and rock into the atmosphere. The ash ejected into the atmosphere created climatic havoc on a global scale. The ash would have blocked sunlight from being received at the ground surface for a few years. A reduction in the reception of solar radiation would have caused the globle climate to cool significantly. Over time this ash settled back to the Earth's surface covering more than half of North America.



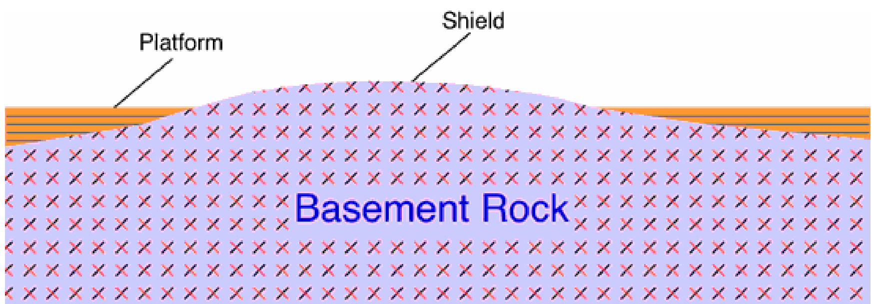
{PRIVATE**}Figure 10n-7**: Map of the location of the Yellowstone caldera. Several large earthquakes have occurred in the last century in the vicinity of the caldera indicating that significant volcanic activity is occurring beneath the ground surface. (**Source:** U.S. Geological Survey - *Yellowstone Volcano Observatory*).

(o) Physiography of the Earth's Terrestrial Surface

{PRIVATE**}**The Earth's ***continental crust*** consists of three discernible units, each with their own characteristics. These units are ***cratons***, **mountain belts**, and ***continental margins***.

Cratons

All of the continental masses on the Earth have a central foundation of very old ***basement rock***. Basement rock is composed of mixtures of ancient ***granite***, ***gneiss***, ***schist***, ***volcanic***, ***plutonic***, and ***sedimentary*** rocks. Some of the Earth's oldest rocks are found in this geologic formation. Basement rock that is exposed at the Earth's surface is called ***shield***. The shields extend for thousands of kilometers and dip ever so slightly from a slightly elevated center. Layers of younger sedimentary strata up to 2000 meters deep cover most of the basement rock. These sedimentary deposits are sometimes called the ***platform*** of the continents. The deposits making up the platform were laid down in shallow seas in repeated episodes over the last 600 million years. The platform and the basement rock together form a ***craton***. The continents of Australia, North America, South America, and Africa each have a single continuous craton forming their nucleus. Eurasia is composed several distinct cratons that are separated from each other by the Alps, Ural, and Himalaya mountain belts.



{PRIVATE**}Figure 10o-1**: Cross section showing the relationship between basement rock and platform sedimentary deposits. Note that the surface of the basement rocks (the shield) is gently arched.

The craton of North America has been relatively stable for about 600 million years. Prior to this period, the North American continent saw several periods of very active growth with the amalgamation of once distinct cratons and the addition of rock along it's margins. Geologic evidence suggests that North America is made up of several once independent minicontinents. Scientists believe that the amalgamation of these minicontinents into the core of the North America continent was complete by about 2.5 billion years ago.

On a global scale, about 70 percent of the Earth's continental crust was formed by 2.5 billion years ago. Over the next 2 billion years, the planet's continents would continue to growth through the ***accretion*** of sedimentary rock and the addition of igneous rocks along the continental margins. This growth was also driven by tectonic processes. The accretion of sedimentary rock occurred with the collision of tectonic plates which pushed ocean sediments onto the continents. Plate subduction created enough heat to melt rock into magma beneath the margins of the continents. This magma then migrated upward through the crust to form ***intrusive*** and ***extrusive igneous*** features and deposits. This process also added significant mass to the continents.

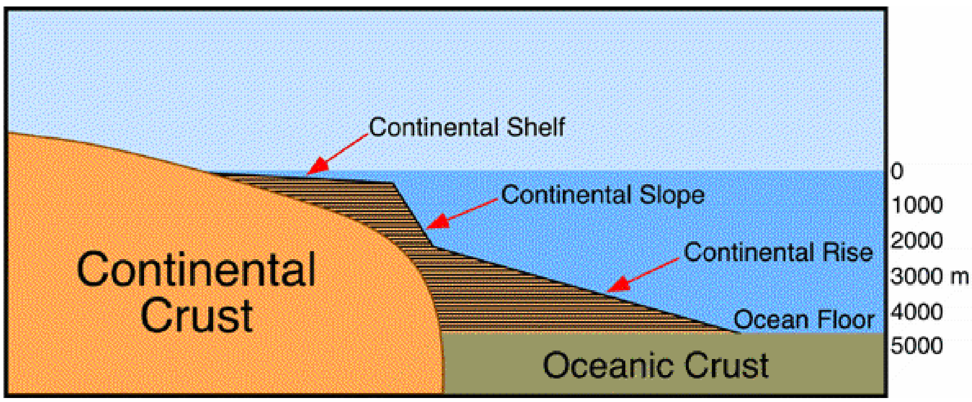
Mountain Belts

Numerous mountain belts are also found on the continents. These features are often located on the edge of cratons. Mountain belts are the result of tectonic processes that cause to crustal plates to collide. This collision results in the ***folding*** and ***faulting*** of rock, igneous intrusive and extrusive activity, and ***metamorphism***. The elevated relief common to mountain belts is generally caused by the compression of rock into a smaller area. Uplift may also be caused by the upward migration of magma through the crust to produce granitic batholiths. Some mountains occur in isolation like ***Mount Rainier*** in the state of Washington, USA. These features are volcanic and are produced by localized extrusive igneous activity.

Continental Margin

Located between the terrestrial continents and the ocean basins is the ***continental margin***. Two basic types of continental margin are recognized: active and passive. Active continental margins occur in the Pacific ocean. Active margins are generally narrow tectonically active areas. They are also associated with earthquakes, oceanic trenches, and volcanoes. Passive continental margins are relatively wide and have a lack of volcanic activity and few earthquakes. The continental margin is actually made up of three structures: the ***continental shelf***, the ***continental slope*** and the ***continental rise***. Both the continental shelf and slope are structurally part of the continents, even though they are below the sea surface.

The ***continental shelf*** is a shallow (average depth 130 meters) gently sloping part of the ***continental crust*** that borders the continents (**Figure 10o-2**). The extent of this feature varies from tens of meters to a maximum width of about 1300 kilometers.



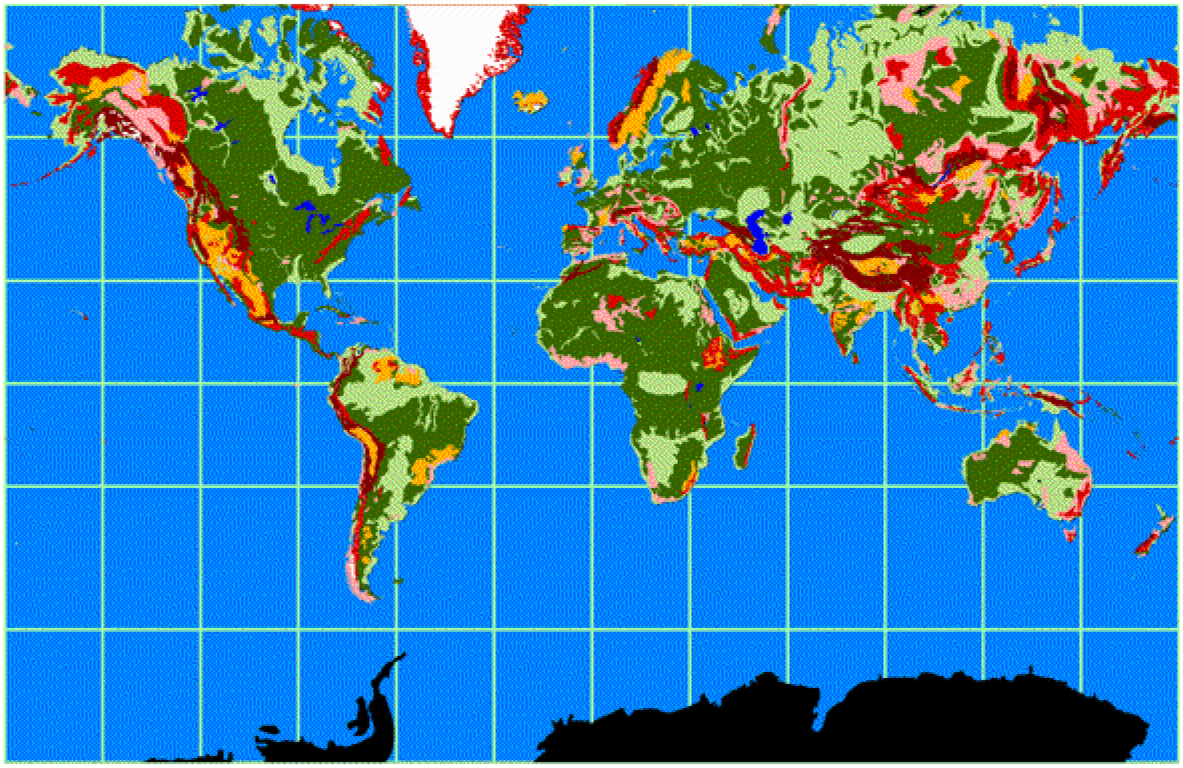
{PRIVATE**}Figure 10o-2**: Marginal features found at the interface of the continents and the ocean basins.

The ***continental slope*** extends from the continental shelf at an average depth of about 135 meters. The base of this steeply sloping (from 1 to 25 degrees, average about 4 degrees) topographic feature occurs at a depth of approximately 2000 meters, marking the edge of the continents. The width of the slope varies from 20 to 100 kilometers. The boundary between the continental slope and shelf is called the ***continental shelf break***.

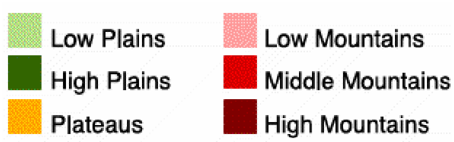
At the base of the continental slope an accumulation of sediments may develop. This accumulation of sediments is properly known as the ***continental rise***. The continental rise is composed of a number ***abyssal fans*** that run side-by-side along the edge of the continental slope. Abyssal fans are usually associated with a deep ***submarine canyon*** cut into the continental slope. Under water abyssal fans can be compared to terrestrial landforms known as ***alluvial fans***. The sediments that make up this feature are transported down the continental slope by turbidity currents, underwater landslides, and several different processes that move clay, silt, and sand. Most of this sediment is terrestrial in origin. The depth of the continental rise ranges from 2000 to 5000 meters deep. The continental rise can be as much as 300 kilometers wide.

Topography of the Terrestrial Surface

Figure 10o-3 classifies the Earth's terrestrial surface in to six different categories based on topography. Most of the Earth's terrestrial surface is dominated by reltively flat low and high plains. The low plains tend to be areas of sediment ***deposition*** because of their low elevation. The high plains can have elevations as high as 600 meters and are more strongly influenced by ***erosion***. Both of these topographic features are often associated with craton's and their exposed shield and platform surfaces. Local relief on both types of plains is less than 100 meters. The three types of mountains shown in **Figure 10o-3** have local relief in excess of 500 meters and slope angles greater than 5 degrees. Many of the "low" mountains are very old structures that have been reduced in height by erosion. Plateaus have altitudes that are greater than the high plains but less than mountains. Local relief of this topographic feature varies between 100 and 500 meters. Some plateaus are the remnants of eroded mountains. Others have formed because of large-scale block faulting.



{PRIVATE**}Figure 10o-3**: The Earth's various topographic regions. The legend below describes the colors associated with the six topographic regions shown. Glaciers are colored white.



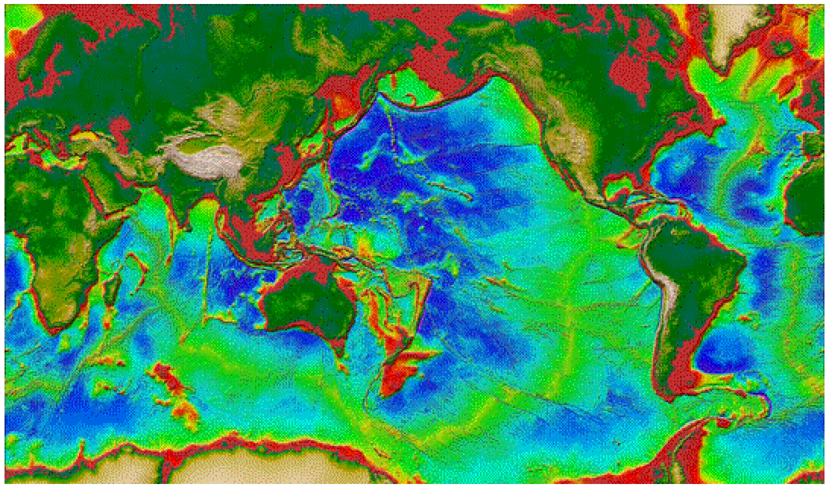
(p) Physiography of the Ocean Basins

{PRIVATE**}Origin of the Ocean Basins**

The ***ocean basins*** are the result of ***tectonic*** forces and processes. All of the ocean basins were formed from volcanic rock that was released from fissures located at the mid-oceanic ridges. The oldest rocks found in these basins are approximately 200 million years old. This is a lot younger than the oldest continental rocks which have ages greater than 4 billion years. The reason for this discrepancy is simple. Tectonic processes destroy old oceanic rocks! Oceanic rock is returned to the Earth's ***mantle*** when ***oceanic crust*** is ***subducted***. Many of these subduction zones occur at the ***continental margins*** where oceanic crust meets ***continental crust***. Subduction also creates the ocean's deep ***trenches***.

Topography of the Ocean Basins

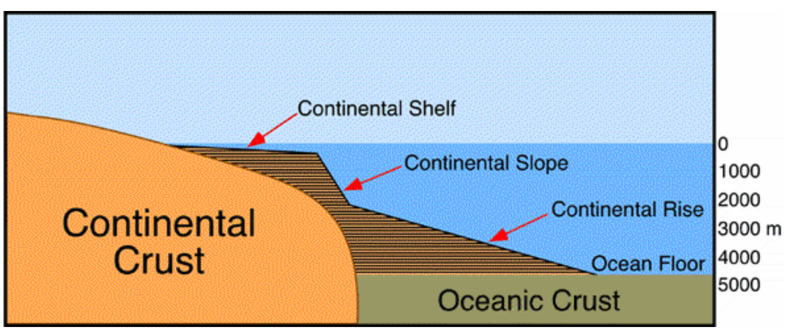
The ocean basins are not featureless Earth surfaces (**Figure 10p-1**). Much of our knowledge about the topographic features that exist here are derived from the following technologies: seismic surveying; echo sounder; side-scan sonar; and the measurement of the height of sea surfaces by satellites. Most of the general information concerning the depth of the ocean basins were made after World War I when the echo sounder was developed for military purposes. This instrument accurately determines the time between the emission of a strong acoustic pulse and the detection of its echo. Using this principle scientists can determine the distance from the sounder to the ocean bottom.



{PRIVATE**}Figure 10p-1**: The following image displays the topography of the Earth's terrestrial land surface and ocean basins. Data for the image comes from satellite altimetry and ship depth soundings, and U.S. Geological Survey digital elevation maps (DEM) of the Earth's land surface. In the ocean basin, the gradation from red to yellow to green to blue indicates increasing depth. A number of topographic features associated with the ocean basin can be seen in this image. The red area that borders the various landmasses is the ***continental shelf***. This feature is structurally part of the continental landmasses despite the fact that it is under water. The yellow to green zone around the continental shelf is the ***continental slope*** and ***continental rise***. The blue region in the various ocean basins constitutes the ***ocean floor***. In the center of ocean basins, the ***mid-oceanic ridges*** can be seen with a color ranging from green to yellow to orange. (**Modified** from image available at the *Seafloor Topography Website*, Institute of Geophysics and Planetary Physics, University of California at San Diego).

Some of the dominant topographic features associated with the ocean basins include:

Continental shelf is a shallow (average depth 130 meters) gently sloping part of the ***continental crust*** that borders the continents (see **Figure 10p-1** and **Figure 10p-2**). The extent of this feature varies from tens of meters to a maximum width of about 1300 kilometers.



{PRIVATE**}Figure 10p-2**: Marginal features found at the interface of the continents and the ocean basins.

The ***continental slope*** extends from the continental shelf at an average depth of about 135 meters (see **Figure 10p-2**). The base of this steeply sloping (from 1 to 25 degrees, average about 4 degrees) topographic feature occurs at a depth of approximately 2000 meters, marking the edge of the continents. The width of the slope varies from 20 to 100 kilometers. Both the continental shelf and slope are considered structurally part of the continents, even though they are below the sea surface. The boundary between the continental slope and shelf is called the ***continental shelf break***.

Submarine canyons are V-shaped canyons cut into the ***continental slope*** to a depth of up to 1200 meters. The submarine canyons are cut perpendicular to the running direction of the continental slope. Many canyons are associated with major rivers such as the Congo, Hudson, and others.

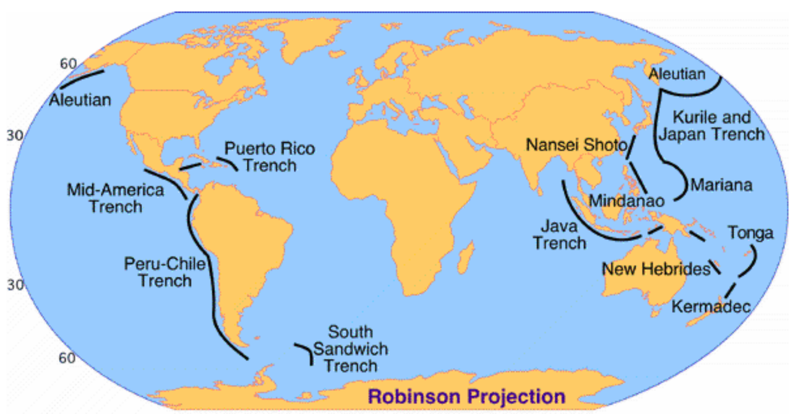
The ***continental rise*** is found at the base of the continental slope (see **Figure 10p-1** and **Figure 10p-2**). The depth of the rise ranges from 2000 to 5000 meters deep. Its breadth is up 300 kilometers wide. This feature was created by the merging of accumulated deposits at the mouths of the many submarine canyons. Each canyon's thick fan-shaped sedimentary deposit is called an ***abyssal fans***.

The ***ocean floor*** is found at the base of the continental rise in water 4000 to 6000 meters deep (see **Figure 10p-1**). The ocean floor accounts for nearly 30 % of the Earth's surface. The composition of the ocean floor consists of a relatively thin layer (on average 5 kilometers thick) of ***basaltic*** rock with an average density of 3.0 grams per cubic centimeter (continents - granite rocks - density 2.7 grams per cubic centimeter).

Numerous ***volcanoes*** populate the floor of the ocean basins. Scientists estimate that there are approximately 10,000 volcanoes on the ocean floor.

Mid-oceanic ridge is normally found rising above the ocean floor at the center of the ocean basins (see **Figure 10p-1**). These features are involved in the generation of new ***oceanic crust*** from volcanic ***fissures*** produced by ***mantle*** up-welling. Some volcanic islands are part of the mid-ocean ridge system (Iceland). The mid-oceanic ridge constitutes 23 % of the Earth's surface. In the center of the mid-oceanic ridge is a rift valley, between 30 to 50 kilometers wide, that dissects 1000 to 3000 meters deep into the ridge system.

Ocean trenches are long, narrow, steep-sided depressions found on the ocean floor that contain the greatest depths in the ocean (11,000 meters - western Pacific). There are 26 oceanic trenches in the world: 3 in the Atlantic Ocean, 1 in the Indian Ocean, and 22 in the Pacific Ocean (**Figure 10p-3**). Generally, the trenches mark the transition between continents and ocean basins, especially in the Pacific basin. Trenches are also the tectonic areas



{PRIVATE**}Figure 10p-3**: Major ocean trenches of the world. The Mariana Trench is the deepest at 11,020 meters below sea-level.

Ocean Basin Configuration

The current spatial configuration of the ***ocean basins*** is a by product of ***plate tectonics***. The creation of new ***oceanic crust*** at the ***mid-oceanic ridge*** moves the continents across the Earth's surface and creates zones of ***subduction***. At the areas of subduction, oceanic crust is forced into the ***mantle*** after it collides with continental crust. Over the past 200 million years, the Atlantic basin has been the most active area of oceanic crust creation. The Atlantic ocean formed about 200 million years ago as the ***Pangaean*** continent began rifting apart. 180 million years ago, North American separated from South America and Africa. North America then joined with Eurasia creating ***Laurasia***. By 135 million years ago, South America began separating from Africa. North America and Eurasia split a few million years after.

11) Introduction to Geomorphology

(a) Models of Landform Development

{PRIVATE**}**The landforms that are found on the surface of the Earth can be grouped into 4 categories:
(1) ***Structural Landforms*** **-** landforms that are created by massive earth movements due to ***plate tectonics***. This includes landforms with some of the following geomorphic features: fold mountains, rift valleys, and volcanoes.
(2) ***Weathering Landforms*** - landforms that are created by the physical or chemical decomposition of rock through ***weathering***. Weathering produces landforms where rocks and sediments are decomposed and disintegrated. This includes landforms with some of the following geomorphic features: karst, patterned ground, and soil profiles.
(3) ***Erosional Landforms*** - landforms formed from the removal of ***weathered*** and ***eroded*** surface materials by wind, water, glaciers, and gravity. This includes landforms with some of the following geomorphic features: river valleys, glacial valleys, and coastal cliffs.
(4) ***Depositional Landforms*** - landforms formed from the ***deposition*** of ***weathered*** and ***eroded*** surface materials. On occasion, these deposits can be compressed, altered by pressure, heat and chemical processes to become ***sedimentary rocks***. This includes landforms with some of the following geomorphic features: beaches, deltas, flood plains, and glacial moraines.
Many landforms show the influence of several of the above processes. We call these landforms ***polygenetic***. Processes acting on landforms can also change over time, and a single landscape can undergo several cycles of development. We call this type landscape development ***polycyclic***.

The following graphical **model** describes the relationship between geomorphic processes and landform types (**Figure 11a-1**).

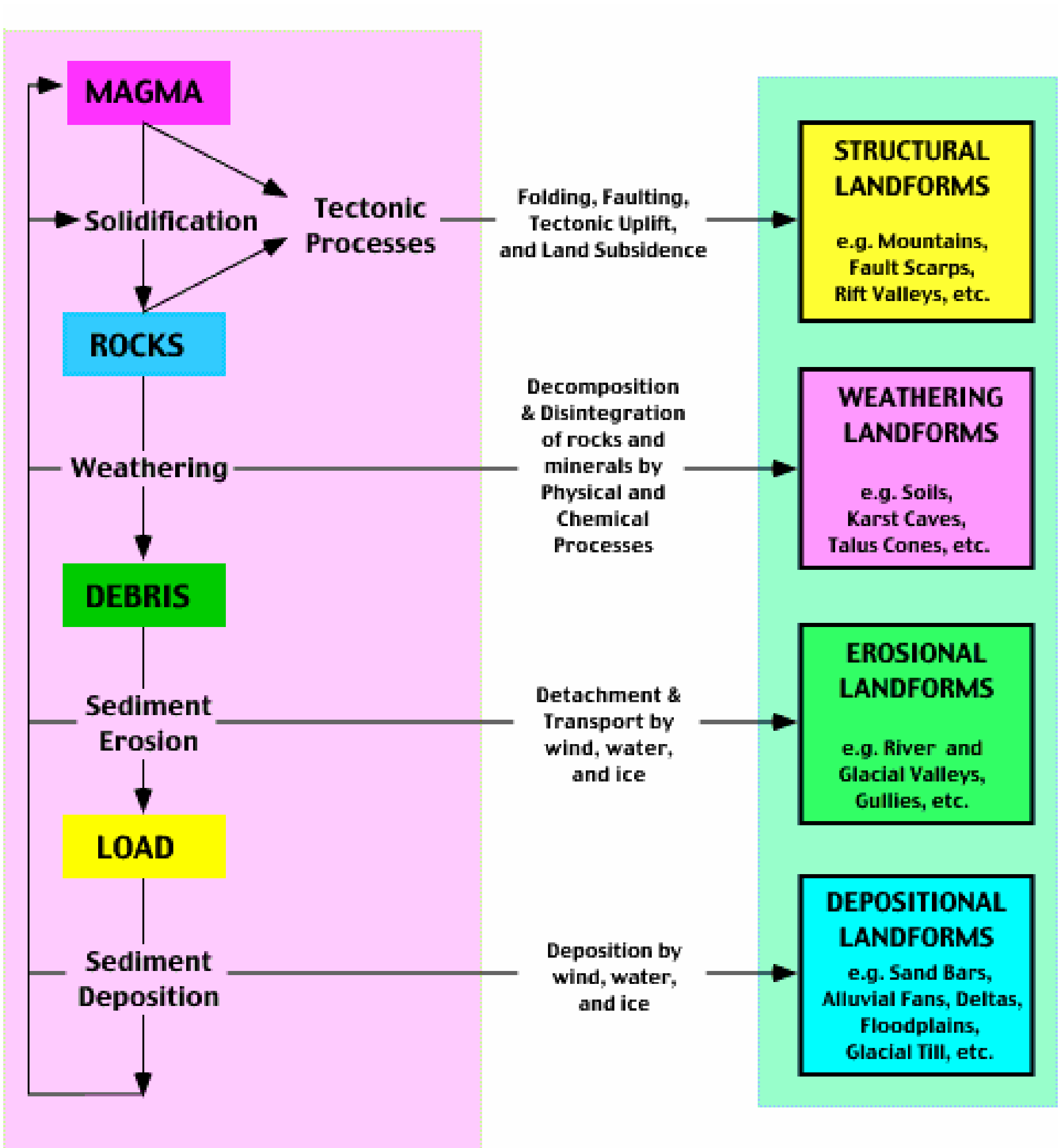


Figure 11a-1: Model of landform development.

(b) Weathering

{PRIVATE}Introduction

Weathering is the breakdown and alteration of rocks and minerals at or near the Earth's surface into products that are more in equilibrium with the conditions found in this environment. Most ***rocks*** and ***minerals*** are formed deep within the Earth's crust where temperatures and pressures differ greatly from the surface. Because the physical and chemical nature of materials formed in the Earth's interior are characteristically in disequilibrium with conditions occurring on the surface. Because of this disequilbrium, these materials are easily attacked, decomposed, and eroded by various chemical and physical surface processes.

Weathering is the first step for a number of other geomorphic and biogeochemical processes. The products of weathering are a major source of sediments for ***erosion*** and ***deposition***. Many types of ***sedimentary rocks*** are composed of particles that have been weathered, eroded, transported, and terminally deposited in basins. Weathering also contributes to the formation of ***soil*** by providing mineral particles like ***sand***, ***silt***, and ***clay***. ***Elements*** and ***compounds*** extracted from the rocks and minerals by weathering processes supply nutrients for plant uptake. The fact that the oceans are saline in the result of the release of ion salts from rock and minerals on the continents. ***Leaching*** and ***runoff*** transport these ions from land to the ocean basins where they accumulate in seawater. In conclusion, weathering is a process that is fundamental to many other aspects of the ***hydrosphere***, ***lithosphere***, and ***biosphere***.

There are three broad categories of mechanisms for ***weathering***: ***chemical***, ***physical*** and ***biological***.

Products of Weathering

The process of weathering can result in the following three outcomes on ***rocks*** and ***minerals***:

(1). The complete loss of particular ***atoms*** or ***compounds*** from the weathered surface.

(2). The addition of specific atoms or compounds to the weathered surface.

(3). A breakdown of one mass into two or more masses, with no chemical change in the mineral or rock.

The residue of weathering consists of chemically altered and unaltered materials. The most common unaltered residue is ***quartz***. Many of the chemically altered products of weathering become very simple small compounds or nutrient ***ions***. These residues can then be dissolved or transported by water, released to the atmosphere as a gas, or taken up by plants for nutrition. Some of the products of weathering, less resistant alumino-silicate minerals, become clay particles. Other altered materials are reconstituted by sedimentary or ***metamorphic*** processes to become new rocks and minerals.

Chemical Weathering

Chemical weathering involves the alteration of the chemical and mineralogical composition of the weathered material. A number of different processes can result in chemical weathering. The most common chemical weathering processes are ***hydrolysis***, ***oxidation***, ***reduction***, ***hydration***, ***carbonation***, and ***solution***.

Hydrolysis is the weathering reaction that occurs when the two surfaces of water and compound meet. It involves the reaction between ***mineral*** ions and the ***ions*** of water (OH- and H+), and results in the decomposition of the ***rock*** surface by forming new compounds, and by increasing the ***pH*** of the solution involved through the release of the hydroxide ions. Hydrolysis is especially effective in the weathering of common silicate and alumino-silicate minerals because of their electrically charged crystal surfaces.

Oxidation is the reaction that occurs between ***compounds*** and oxygen. The net result of this reaction is the removal of one or more electrons from a compound, which causes the structure to be less rigid and increasingly unstable. The most common oxides are those of iron and aluminum, and their respective red and yellow staining of soils is quite common in tropical regions which have high temperatures and precipitation. ***Reduction*** is simply the reverse of oxidation, and is thus caused by the addition of one or more electrons producing a more stable compound.

Hydration involves the rigid attachment of H+ and OH- ***ions*** to a reacted compound. In many situations the H and OH ions become a structural part of the crystal lattice of the mineral. Hydration also allows for the acceleration of other decompositional reactions by expanding the crystal lattice offering more surface area for reaction.

Carbonation is the reaction of carbonate and bicarbonate ***ions*** with minerals. The formation of carbonates usually takes place as a result of other chemical processes. Carbonation is especially active when the reaction environment is abundant with carbon dioxide. The formation of carbonic acid, a product of carbon dioxide and water, is important in the solution of carbonates and the decomposition of mineral surfaces because of its acidic nature.

Water and the ions it carries as it moves through and around rocks and minerals can further the weathering process. Geomorphologists call this phenomena ***solution***. The effects of dissolved carbon dioxide and hydrogen ions in water have already been mentioned, but solution also entails the effects of a number of other dissolved compounds on a mineral or rock surface. Molecules can mix in solution to form a great variety of basic and acidic decompositional compounds. The extent, however, of rock being subjected to solution is determined primarily by climatic conditions. Solution tends to be most effective in areas that have humid and hot climates.

The most important factor affecting all of the above mentioned chemical weathering processes is climate. Climatic conditions control the rate of weathering that takes place by regulating the catalysts of moisture and temperature. Experimentation has discovered that tropical weathering rates, where temperature and moisture are at their maximum, are three and a half times higher than rates in temperate environments.

Physical Weathering

Physical weathering is the breakdown of mineral or rock material by entirely mechanical methods brought about by a variety of causes. Some of the forces originate within the rock or mineral, while others are applied externally. Both of these stresses lead to strain and the rupture of the rock. The processes that may cause mechanical rupture are ***abrasion***, **crystallization**, **thermal insolation**, ***wetting and drying***, and **pressure release**.

Abrasion occurs when some force causes two rock surfaces to come together causing mechanical wearing or grinding of their surfaces. Collision between rock surfaces normally occurs through the ***erosional*** transport of material by wind, water, or ice.

Crystallization can cause the necessary stresses needed for the mechanical rupturing of rocks and minerals. Crystal growth causes stress as a result of a compound's or an element's change of physical state with change in temperature. The transformation from liquid to solid crystalline form produces a volumetric change which in turn causes the necessary mechanical action for rupture. There are primarily two types of crystal growth that occur; they are ice and salt. Upon freezing the volumetric change of water from liquid to solid is 9 %. This relatively large volumetric change upon freezing has potentially a great rupturing effect. Several researchers have discovered in the laboratory and the field that frost action plays a major role in weathering in temperate and polar regions of the Earth. The threshold temperature for frost action is at least - 5 degrees Celsius, and it is at this temperature that the most effective rupturing occurs.

The crystallization of salt exhibits volumetric changes from 1 to 5 percent depending on the temperature of the rock or mineral surface. Most salt weathering occurs in hot arid regions, but it may also occur in cold climates. For example, cavernous salt weathering of granite is widespread in the dry valley regions of South Victoria Land, Antarctica. At this location outcrops and large boulders are pitted by holes up to 2 meters in diameter. Researchers have also found that frost weathering is greatly enhanced by the presence of salt.

The physical breakdown of rock by their expansion and contraction due to diurnal temperature changes is one of the most keenly debated topics in rock weathering research. Known as ***insolation weathering***, it is the result of the physical inability of rocks to conduct heat well. This inability to conduct heat results in differential rates of expansion and contraction. Thus, the surface of the rock expands more than its interior, and this stress will eventually cause the rock to rupture. Differential expansion and contraction may also be due to the variance in the colors of mineral grains in rock. Dark colored grains, because of their absorptive properties, will expand much more than light colored grains. Therefore, in a rock peppered with many different colored grains, rupturing can occur at different rates at the various mineral boundaries.

Alternate ***wetting and drying*** of rocks, sometimes known as ***slaking***, can be a very important factor in weathering. Slaking occurs by the mechanism of "ordered water", which is the accumulation of successive layers of water molecules in between the mineral grains of a rock. The increasing thickness of the water pulls the rock grains apart with great tensional stress. Recent research has shown that slaking in combination with dissolved sodium sulfate can disintegrate samples of rock in only twenty cycles of wetting and drying.

Pressure release of rock can cause physical weathering due to ***unloading***. The majority of igneous rocks were created deep under the Earth's surface at much higher pressures and temperatures. As ***erosion*** brings these rock formations to the surface, they become subjected to less and less pressure. This unloading of pressure causes the rocks to fracture horizontally with an increasing number of fractures as the rock approaches the Earth's surface. Spalling, the vertical development of fractures, occurs because of the bending stresses of unloaded sheets across a three dimensional plane.

Biological Weathering

Biological weathering involves the disintegration of rock and mineral due to the chemical and/or physical agents of an organism. The types of organisms that can cause weathering range from bacteria to plants to animals.

Biological weathering involves processes that can be either chemical or physical in character. Some of the more important processes are:

1. Simple breaking of particles, by the consumption of soils particles by animals. Particles can also fracture because of animal burrowing or by the pressure put forth by growing roots.

2. Movement and mixing of materials. Many large soil organisms cause the movement of soil particles. This movement can introduce the materials to different weathering processes found at distinct locations in the soil profile.

3. Simple chemical processes like ***solution*** can be enhanced by the carbon dioxide produced by ***respiration***. Carbon dioxide mixing with water forms carbonic acid.

4. The complex chemical effects that occur as a result of ***chelation***. Chelation is a biological process where organisms produce organic substances, known as ***chelates***, that have the ability to decompose minerals and rocks by the removal of metallic ***cations***.

5. Organisms can influence the moisture regime in soils and therefore enhance weathering. Shade from aerial leaves and stems, the presence of roots masses, and ***humus*** all act to increase the availability of water in the soil profile. Water is a necessary component in several physical and chemical weathering processes.

6. Organisms can influence the ***pH*** of the soil solution. ***Respiration*** from plant roots releases carbon dioxide. If the carbon dioxide mixes with water carbonic acid is formed which lowers soil pH. ***Cation exchange*** reactions by which plants absorb nutrients from the soil can also cause pH changes. The absorption processes often involves the exchange of ***basic*** cations for hydrogen ions. Generally, the higher the concentration of hydrogen ions the more ***acidic*** a soil becomes.

(c) Landforms of Weathering

{PRIVATE}Regolith and Soil

Most landforms to some extent show the effects of ***weathering***. On the bedrock surface of these landscapes are the accumulations of the products of weathering. Within these accumulations are materials displaying various degrees of physical, chemical, and biological alteration. These materials range in size from large boulders to clay sized particles less than 0.004 millimeters in diameter. Geomorphologists refer to these accumulations as ***regolith***. Regolith can be further altered by climate, organisms, and topography over time to create ***soil***. Soil is the most obvious landform of weathering.

Limestone Landforms

Among the most interesting and most beautiful landforms of weathering are those which develop in regions of ***limestone*** bedrock. These landscapes are commonly called ***karst***. In karst landscapes weathering is concentrated along joints and bedding planes of the limestone producing a number of different sculptured features from the effects of ***solution***. Depressions of all sizes and shapes pit the landscape surface and are the most obvious features associated with karst. Beneath the surface, solution results in the formation of caves, ***springs***, underground water channels, and deposits from ***evaporation***.

Periglacial Landforms

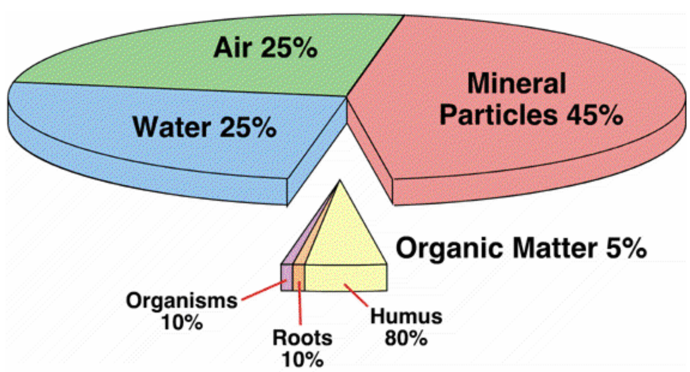
Unique weathering landforms are also found in pola r and sub-polar regions. In these regions, physical weathering processes are dominant, with active freeze-thaw and frost-shattering being the most active. Associated with these weathering processes are a number of unique surface features that develop only in ***periglacial*** environments. Collectively known as ***patterned ground***, these surface features resemble circles, polygons, nets, steps, and stripes. The outlines of all of these features consist of elevated accumulations of coarse regolith fragments. Scientists believe that these outlines result from the systematic sorting of particles of a wide range of texture sizes by freeze-thaw action. The sorting causes larger fragments to move vertically upward and horizontally outward. Horizontal movement stops when one feature encounters another, linking the perimeter of two or more features. The linking of many adjacent features creates net-like patterns.

(d) Introduction to Soils

{PRIVATE**}Introduction**

An important factor influencing the productivity of any ***ecosystem*** is the nature of its ***soils***. Soils are vital for the existence of many forms of life that have evolved on our planet. For example, soils provide vascular plants with a medium for growth and supply these organisms with most of their nutritional requirements. Plants are the basis of virtually all terrestrial ***food chains***.

Soil itself is a much more complex phenomenon than most people realize. It is certainly not just fine ***mineral*** particles or dirt . A true soil also contains air, water, and organic matter (**Figure 11d-1**). The formation of a soil is influenced by organisms, climate, topography, parent material, and time. The following items describe some important features of a soil that help to distinguish it from mineral sediments.



{PRIVATE**}Figure 11d-1:** Most soils contain four basic components: mineral particles, water, air, and organic matter. Organic matter can be further sub-divided into humus, roots, and living organisms. The values given above are for an average soil.

Organic Activity

A mass of mineral particles alone do not constitute a true ***soil***. True soils are influenced, modified, and supplemented by living organisms. Plants and animals aid in the development of a soil through the addition of

organic matter. Fungi and bacteria reduce this organic matter to a semi-soluble chemical complex called ***humus***. Larger soil organisms, like earthworms, beetles, and termites, mix humus into the mineral matter of the soil.

Humus is the biochemical substance that makes the upper layers of the soil become dark. Humus is itself colored dark brown to black. Humus is difficult to see or study in isolation because it becomes intimately mixed with mineral particles. Humus provides the soil with a number of benefits:

It increases the soil's ability to hold and store moisture.

It reduces the ***eluviation*** of soluble nutrients.

It is an important source of carbon and nitrogen required by plants.

It improves soil structure for plant growth.

Organic activity in abundant in soils. One cubic centimeter of soil may contain over 1,000,000 ***bacteria***. A hectare of pasture land in a humid climate can contain more than a million earthworms and about 25 million insects. Insects and earthworms are very important in mixing and aerating the soil. These organisms are also responsible for producing a significant part of a soil's humus through the incomplete digestion of organic matter.

Translocation

When water moves downward into the ***soil***, it causes both mechanical and chemical translocations of material. The process of ***eluviation*** is the flushing of fine particles (like ***clay***) or dissolved substances to lower levels in the ***soil profile***. The process of deposition of these fine particles at the lower level is called ***illuviation***. Downward percolating water also causes the translocation of nutrients and complex chemical substances within the soil profile. Like the finest particles, these dissolved substances are removed from the surface layer and are either removed by groundwater outflow or deposited at a lower level in the soil. The complete chemical removal of these substances from the soil profile is known as ***leaching***.

Soil Texture

The ***texture*** of a ***soil*** refers to the size distribution of the mineral particles composing the soil. Particles are normally grouped into three main classes: ***sand***, ***silt***, and ***clay***. **Table 11d-1** describes the classification of soil particles according to size.

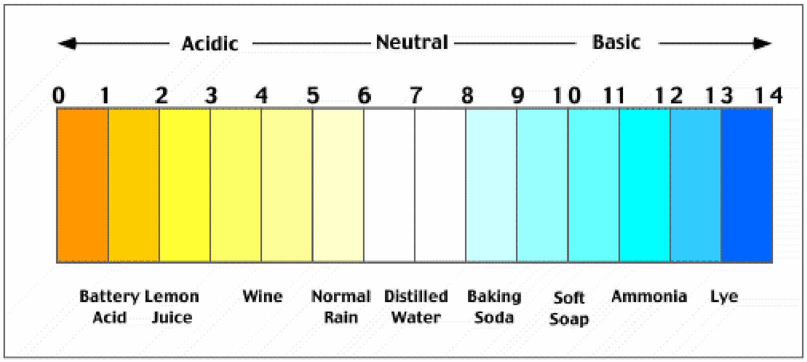
Table 11d-1: Particle size ranges for sand, silt, and clay.

{PRIVATE}**Type of Mineral Particle**	**Size Range**
Sand	2.0 - 0.06 millimeters
Silt	0.06 - 0.002 millimeters
Clay	less than 0.002 millimeters

Of all the constituents of the mineral portion of the soil, ***clay*** is probably the most important. Despite their small size, clay particles have a very large surface area relative to their volume. This large surface is highly reactive and has the ability to attract and hold positively charged nutrient ions. These nutrients are available to plant roots for nutrition. Clay particles are also somewhat flexible and plastic because of their lattice-like design. This feature allows clay particles to absorb water and other substances into their structure.

Soil pH

Soils support a number of inorganic and organic chemical reactions. Many of these reactions are dependent on a variety of soil chemical properties. One of the most important chemical properties of a soil is ***pH*** (**Figure 11d-2**). Soil pH is generally related to the concentration of free hydrogen ions in the soil matrix. Hydrogen ions are made available to the soil matrix by the ***dissociation*** of water, by the chemical activity of roots, and by many ***chemical weathering*** reactions. The concentration of hydrogen ions determines the pH of the soil. Soils with a relatively large concentration of hydrogen ions tend to be ***acidic***. ***Alkaline*** soils have a relatively low concentration of hydrogen ions.



{PRIVATE}**Figure 11d-2:** The pH scale. A value of 7.0 is considered neutral. Values higher than 7.0 are increasingly alkaline or basic. Values lower than 7.0 are increasingly acidic. The illustration above also describes the pH of some common substances.

Soil fertility is directly influenced by pH through the solubility of many nutrients. At a pH lower than 5.5, many nutrients become very soluble and are readily leached from the soil profile. At high pH, nutrients become insoluble and plants cannot readily extract them. Maximum soil fertility occurs in the range 6.0 to 7.2.

Soil Color

Soils tend to have distinct variations in color both horizontally and vertically. The coloring of soils occurs because of a variety of factors. Soils of the humid tropics are generally red or yellow because of the oxidation of iron or aluminum, respectively. In the temperate grasslands, large additions of humus cause soils to be black. The heavy leaching of iron causes coniferous forest soils to be gray. High water tables in soils cause the reduction of iron, and these soils tend to have greenish and gray-blue hues. Organic matter colors the soil black. The combination of iron oxides and organic content gives many soil types a brown color. Other coloring materials sometimes present include white calcium carbonate, black manganese oxides, and black carbon compounds.

Soil Profiles

Most ***soils*** have a distinctive ***profile*** or sequence of horizontal layers. Generally, these horizons result from the soil processes of ***eluviation*** and ***organic*** activity. Five general layers are normally present in a typical soil: **O**, **A**, **B**, **C**, and **R horizons** (**Figure 11d-3**).

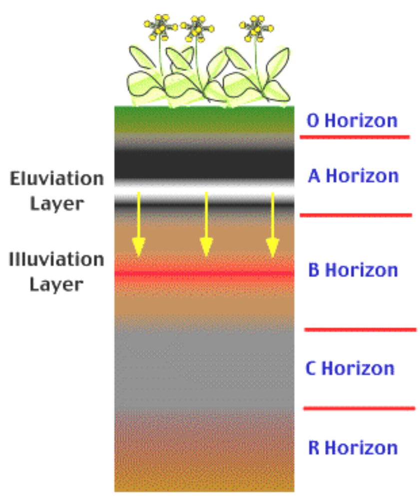


Figure 11d-3: Typical layers found in a soil profile.

The ***O horizon*** is the topmost layer of most soils. It is composed mainly of plant ***litter*** at various levels of decomposition and ***humus***.

Below it is the ***A horizon***. This layer is composed primarily of ***mineral*** particles. which has two characteristics: it is the layer in which humus and other organic materials are mixed with mineral particles, and it is a zone of translocation from which eluviation has removed finer particles and soluble substances, both of which may be deposited at a lower layer. Thus the A horizon is dark in color and usually light in texture and porous. The A horizon is commonly differentiated into a darker upper horizon or organic accumulation, and a lower horizon showing loss of material by eluviation.

The ***B horizon*** is a mineral soil layer which is dominated by ***illuviation***. It receives material eluviated from the A horizon. This layer also has a higher bulk density than the A horizon due to its enrichment of clay particles. The B horizon may be colored by oxides of iron and aluminum or by calcium carbonate illuviated from the A horizon.

The ***C horizon*** is composed of ***weathered*** parent material that has not been yet significantly affected by the ***pedogenic*** processes or translocation and organic modification.

The ***R horizon*** consists of unweathered bedrock.

(e) Soil Pedogenesis

Pedogenesis can be defined as the process of soil development. Late in the 19th century, scientists Hilgard in the United States and the Russian Dukuchaev both suggested independently that pedogenesis was principally controlled by climate and vegetation. This idea was based on the observation that comparable soils developed in spatially separate areas when their climate and vegetation were similar. In the 1940s, Hans Jenny extended these ideas based on the observations of many subsequent studies examining the processes involved in the formation of soils. Jenny believed that the kinds of soils that develop in a particular area are largely determined by five interrelated factors: **climate**; **living organisms**; **parent material**; **topography**; and **time**.

Climate plays a very important role in the genesis of a soil. On the global scale, there is an obvious correlation between major soil types and the Köppen climatic classification systems major climatic types. At regional and local scales, climate becomes less important in soil formation. Instead, pedogenesis is more influenced by factors like parent material, topography, vegetation, and time. The two most important climatic variables influencing soil formation are temperature and moisture. ***Temperature*** has a direct influence on the ***weathering*** of bedrock to produce mineral particles. Rates of bedrock weathering generally increase with higher temperatures. Temperature also influences the activity of soil microorganisms, the frequency and magnitude of soil chemical reactions, and the rate of plant growth. Moisture levels in most soils are primarily controlled by the addition of water via ***precipitation*** minus the losses due to ***evapotranspiration***. If additions of water from precipitation surpass losses from evapotranspiration, moisture levels in a soil tend to be high. If the water loss due to evapotranspiration exceeds inputs from precipitation, moisture levels in a soil tend to be low. High moisture availability in a soil promotes the weathering of bedrock and sediments, chemical reactions, and plant growth. The availability of moisture also has an influence on soil ***pH*** and the ***decomposition*** of ***organic matter***.

Living Organisms have a role in a number of processes involved in pedogenesis including organic matter accumulation, profile mixing, and ***biogeochemical nutrient cycling***. Under equilibrium conditions, vegetation and soil are closely linked with each other through nutrient cycling. The cycling of nitrogen and carbon in soils is almost completely controlled by the presence of animals and plants. Through ***litterfall*** and the process of ***decomposition***, organisms add ***humus*** and nutrients to the soil which influences soil structure and fertility. Surface vegetation also protects the upper layers of a soil from ***erosion*** by way of binding the soils surface and reducing the speed of moving wind and water across the ground surface.

Parent Material refers to the ***rock*** and ***mineral*** materials from which the soils develop. These materials can be derived from residual sediment due to the weathering of bedrock or from sediment transported into an area by way of the erosive forces of wind, water, or ice. Pedogenesis is often faster on transported sediments because the weathering of parent material usually takes a long period of time. The influence of parent material on pedogenesis is usually related to soil texture, soil chemistry, and nutrient cycling.

Topography generally modifies the development of soil on a local or regional scale. Pedogenesis is primarily influenced by topography's effect on microclimate and drainage. Soils developing on moderate to gentle slopes are often better drained than soils found at the bottom of valleys. Good drainage enhances an number of pedogenic processes of ***illuviation*** and ***eluviation*** that are responsible for the development of soil horizons. Under conditions of poor drainage, soils tend to be immature. Steep topographic gradients inhibit the development of soils because of erosion. Erosion can retard the development through the continued removal of surface sediments. Soil microclimate is also influenced by topography. In the Northern Hemisphere, south facing slopes tend to be warmer and drier than north facing slopes. This difference results in the soils of the two areas being different in terms of depth, texture, biological activity, and ***soil profile*** development.

Time influences the temporal consequences of all of the factors described above. Many soil processes become ***steady state*** overtime when a soil reaches maturity. Pedogenic processes in young soils are usually under active modification through ***negative*** and ***positive feedback*** mechanisms in attempt to achieve ***equilibrium***.

Principal Pedogenic Processes

A large number of processes are responsible for the formation of soils. This fact is evident by the large number of different types of soils that have been classified by soil scientists (see topic *11f*). However, at the macro-scale we can suggest that there are five main principal pedogenic processes acting on soils. These processes are **laterization**, **podzolization**, **calcification**, **salinization**, and **gleization**.

Laterization is a pedogenic process common to soils found in tropical and subtropical environments. High temperatures and heavy precipitation result in the rapid ***weathering*** of rocks and minerals. Movements of large amounts of water through the soil cause ***eluviation*** and ***leaching*** to occur. Almost all of the by products of weathering, very simple small compounds or nutrient ***ions***, are translocated out of the soil profile by leaching if not taken up by plants for nutrition. The two exceptions to this process are iron and aluminum compounds. Iron oxides give tropical soils their unique reddish coloring. Heavy leaching also causes these soils to have an ***acidic pH*** because of the net loss of ***base cations***.

Podzolization is associated with humid cold mid-latitude climates and ***coniferous vegetation***. Decomposition of coniferous litter and heavy summer precipitation create a soil solution that is strongly acidic. This ***acidic soil solution*** enhances the processes of ***eluviation*** and leaching causing the removal of soluble ***base cations*** and aluminum and iron compounds from the ***A horizon***. This process creates a sub-layer in the A horizon that is white to gray in color and composed of silica sand.

Calcification occurs when evapotranspiration exceeds precipitation causing the upward movement of dissolved alkaline salts from the groundwater. At the same time, the movement of rain water causes a downward movement of the salts. The net result is the deposition of the translocated cations in the ***B horizon***. In some cases, these deposits can form a hard layer called ***caliche***. The most common substance involved in this process is ***calcium carbonate***. Calcification is common in the prairie ***grasslands***.

Salinization is a process that functions in the similar way to calcification. It differs from calcification in that the salt deposits occur at or very near the soil surface. Salinization also takes place in much drier climates.

Gleization is a pedogenic process associated with poor drainage. This process involves the accumulations of organic matter in the upper layers of the soil. In lower horizons, mineral layers are stained blue-gray because of the chemical ***reduction*** of iron.

(f) Soil Classification

Soil Classification Systems have been developed to provide scientists and resource managers with generalized information about the nature of a soil found in a particular location. In general, environments that share comparable soil forming factors produce similar types of soils. This phenomenon makes classification possible. Numerous classification systems are in use worldwide. We will examine the systems commonly used in the United States and Canada.

United States Soil Classification System

The first formal system of soil classification was introduced in the United States by Curtis F. Marbut in the 1930s. This system, however, had some serious limitations, and by the early 1950s the **United States Soil Conservation Service** began the development of a new method of soil classification. The process of development of the new system took nearly a decade to complete. By 1960, the review process was completed and the **Seventh Approximation Soil Classification System** was introduced. Since 1960, this soil classification system has undergone numerous minor modifications and is now under the control of **Natural Resources Conservation Service** (**NRCS**), which is a branch of the **Department of Agriculture**. The current version of the system has six levels of classification in its hierarchical structure. The major divisions in this classification system, from general to specific, are: orders, suborders, great groups, subgroups, families, and series. At its lowest level of organization, the U.S. system of soil classification recognizes approximately 15,000 different soil series.

The most general category of the **NRCS Soil Classification System** recognizes eleven distinct soil orders: **oxisols**, **aridsols**, **mollisols**, **alfisols**, **ultisols**, **spodsols**, **entisols**, **inceptisols**, **vertisols**, **histosols**, and **andisols**.

Oxisols develop in tropical and subtropical latitudes that experience an environment with high precipitation and temperature. The profiles of oxisols contain mixtures of quartz, kaolin clay, iron and aluminum oxides, and organic matter. For the most part they have a nearly featureless soil profile without clearly marked horizons. The abundance of iron and aluminum oxides found in these soils results from strong chemical ***weathering*** and heavy ***leaching***. Many oxisols contain ***laterite*** layers because of a seasonally fluctuating ***water table***.

Aridsols are soils that develop in very dry environments. The main characteristic of this soil is poor and shallow soil horizon development. Aridsols also tend to be light colored because of limited humus additions from vegetation. The hot climate under which these soils develop tends to restrict vegetation growth. Because of limited rain and high temperatures soil water tends to migrate in these soils in an upward direction. This condition causes the deposition of salts carried by the water at or near the ground surface because of evaporation. This soil process is of course called ***salinization***.

Mollisols are soils common to grassland environments. In the United States most of the natural grasslands have been converted into agricultural fields for crop growth. Mollisols have a dark colored surface horizon, tend to be base rich, and are quite fertile. The dark color of the ***A horizon*** is the result of ***humus*** enrichment from the decomposition of ***litterfall***. Mollisols found in more arid environments often exhibit ***calcification***.

Alfisols form under forest vegetation where the parent material has undergone significant weathering. These soils are quite widespread in their distribution and are found from southern Florida to northern Minnesota. The most distinguishing characteristics of this soil type are the ***illuviation*** of ***clay*** in the ***B horizon***, moderate to high concentrations of base cations, and light colored surface horizons.

Ultisols are soils common to the southeastern United States. This region receives high amounts of precipitation because of summer ***thunderstorms*** and the winter dominance of the ***mid-latitude cyclone***. Warm temperatures and the abundant availability of moisture enhances the weathering process and increases the rate of ***leaching*** in these soils. Enhanced weathering causes mineral alteration and the dominance of iron and aluminum oxides. The presence of the iron oxides causes the ***A horizon*** of these soils to be stained red. Leaching causes these soils to have low quantities of ***base cations***.

Spodsols are soils that develop under ***coniferous vegetation*** and as a result are modified by ***podzolization***. Parent materials of these soils tend to be rich in ***sand***. The litter of the coniferous vegetation is low in base cations and contributes to acid accumulations in the soil. In these soils, mixtures of organic matter and aluminum, with or without iron, accumulate in the ***B horizon***. The ***A horizon*** of these soils normally has an eluvial layer that has the color of more or less quartz sand. Most spodosols have little ***silicate clay*** and only small quantities of ***humus*** in their ***A horizon***.

Entisols are immature soils that lack the vertical development of horizons. These soils are often associated with recently deposited sediments from wind, water, or ice erosion. Given more time, these soils will develop into another soil type.

Inceptisols are young soils that are more developed than entisols. These soils are found in arctic tundra environments, glacial deposits, and relatively recent deposits of stream ***alluvium***. Common characteristics of recognition include immature development of ***eluviation*** in the ***A horizon*** and ***illuviation*** in the ***B horizon***, and evidence of the beginning of ***weathering*** processes on parent material sediments.

Vertisols are heavy ***clay*** soils that show significant expansion and contraction due to the presence or absence of moisture. Vertisols are common in areas that have shale parent material and heavy precipitation. The location of these soils in the United States is primarily found in Texas where they are used to grow cotton.

Histosols are organic soils that form in areas of poor drainage. Their profile consists of thick accumulations of ***organic matter*** at various stages of ***decomposition***.

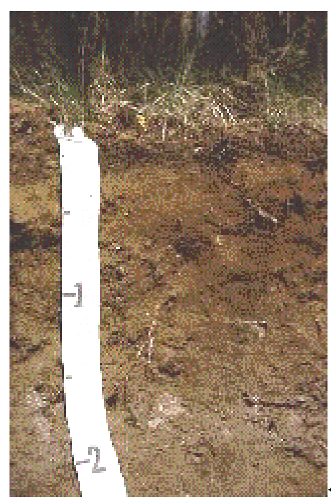
Andisols develop from volcanic parent materials. Volcanic deposits have a unique process of ***weathering*** that causes the accumulation of ***allophane*** and oxides of iron and aluminum in developing soils.

Canadian System of Soil Classification

Canada's first independent taxonomic system of soil classification was first introduced in 1955. Prior to 1955, systems of classification used in Canada were strongly based on methods being applied in the United States. However, the U.S. system was based on environmental conditions common to the United States. Canadian soil scientists required a new method of soil classification that focused on pedogenic processes in cool climatic environments.

Like the US system, the **Canadian System of Soil Classification** differentiates soil types on the basis of measured properties of the profile and uses a hierarchical scheme to classify soils from general to specific. The most recent version of the classification system has five categories in its hierarchical structure. From general to specific, the major categories in this system are: orders, great groups, subgroups, families, and series. At its most general level, the **Canadian System** recognizes nine different **soil orders**:

(1) **Brunisol** - is a normally immature soil commonly found under forested ecosystems. The most identifying trait of these soils is the presence of a ***B horizon*** that is brownish in color. The soils under the dry pine forests of south-central British Columbia are typically brunisols.



{PRIVATE**}**

Brunisolic Pine Landscape (Central British Columbia)   Brunisol Profile

Figure 11f-1: Associated surface environment and profile of a brunisol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(2) **Chernozem** - is a soil common to grassland ecosystems. This soil is dark in color (brown to black) and has an ***A horizon*** that is rich in ***organic matter***. Chernozems are common in the Canadian prairies. The images below are from the eastern prairies where higher seasonal rainfalls produce black chernozemic soils.

{PRIVA TE}

Chernozemic Landscape (Prairies)

Chernozen Profile

Figure 11f-2: Associated surface environment and profile of a chernozem soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(3) **Cryosol** - is a high latitudes soil common in the tundra. This soil has a layer of ***permafrost*** within one meter of the soil surface. The image on the left is of tundra landscape dominated by moss and lichen vegetation. The soil profile has a permanently frozen ice wedge beneath its surface.

{PRIVATE}

Tundra Cryosolic Landscape (N.W.T.)

Organic Cryosol Profile

{PRIVATE}**Figure 11f-3:** Associated surface environment and profile of a cryosol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(4) **Gleysol** - is a soil found in an ecosystem that is frequently flooded or permanently waterlogged. Its soil horizons show the chemical signs of oxidation and reduction.

{PRIVATE}

Flooded Gleysolic Landscape (Atlantic Coast) Gleysol Profile

{PRIVATE}**Figure 11f-4:** Associated surface environment and profile of a gleysol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(5) **Luvisol** - is another type of soil that develops under forested conditions. This soil, however, has a calcareous parent material which results in a high ***pH*** and strong ***eluviation*** of ***clay*** from the ***A horizon***.

{PRIVATE}

Luvisolic Sub-Boreal Forest Landscape (Northern British Columbia) Luvisol Profile

{PRIVATE}**Figure 11f-5:** Associated surface environment and profile of a luvisol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(6) **Organic** - this soil is mainly composed of ***organic matter*** in various stages of ***decomposition***. Organic soils are common in fens and bogs. The profiles of these soils have an obvious absence of mineral soil particles.

{PRIVATE}

Organic Soil Landscape (British Columbia) Organic Soil Profile

{PRIVATE}**Figure 11f-6:** Associated surface environment and profile of an organic soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

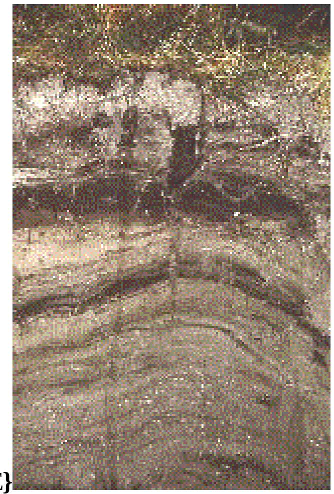
(7) **Podzol** - is a soil commonly found under coniferous forests. Its main identifying traits are a poorly decomposed organic layer, an ***eluviated*** A horizon, and a B horizon with ***illuviated organic matter***, aluminum, and iron. The forested regions of southern Ontario and the temperate rainforests of British Columbia normally have podzolic soils.

{PRIVATE}

Forested Podzolic Landscape (Ontario) Podzol Profile

{PRIVATE}**Figure 11f-7:** Associated surface environment and profile of a podzol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

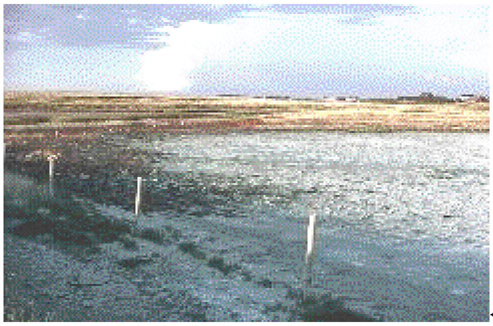
(8) **Regosol** - is any young underdeveloped soil. Immature soils are common in geomorphically dynamic environments. Many mountain river valleys in British Columbia have floodplains with surface deposits that are less than 3000 years old. The soils in these environments tend to be regosols.

{PRIVATE}

Immature Regosolic Landscape (Floodplain British Columbia) Regosol Profile

{PRIVATE}**Figure 11f-8:** Associated surface environment and profile of a regosol soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(9) **Solonetzic** - is a grassland soil where high levels of ***evapotranspiration*** cause the deposition of salts at or near the soil surface. Solonetzic soils are common in the dry regions of the prairies where evapotranspiration greatly exceeds precipitation input. The movement of water to the earth's surface because of capillary action, transpiration, and evaporation causes the deposition of salts when the water evaporates into the atmosphere.

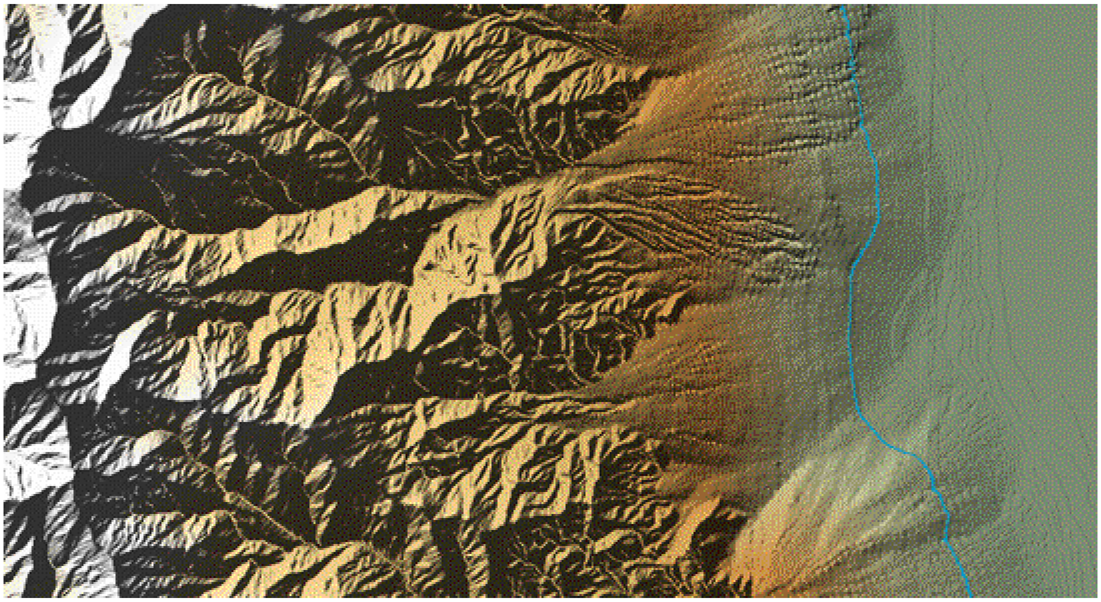
{PRIVATE}

Saline Solonetzic Landscape (Saskatchewan) Solonetzic Profile

{PRIVATE}**Figure 11f-9:** Associated surface environment and profile of a solonetzic soil. (Images from *Soil Landscapes of Canada*, Version 2.2, Agriculture and Agri-Food Canada. 1996.)

(g) Erosion and Deposition

Erosion is defined as the removal of ***soil***, ***sediment***, ***regolith***, and ***rock*** fragments from the landscape. Most landscapes show obvious evidence of erosion. Erosion is responsible for the creation of hills and valleys. It removes sediments from areas that were once glaciated, shapes the shorelines of lakes and coastlines, and transports material downslope from elevated sites. In order for erosion to occur three processes must take place: ***detachment***, ***entrainment*** and ***transport***. Erosion also requires a medium to move material. Wind, water, and ice are the mediums primarily responsible for erosion. Finally, the process of erosion stops when the transported particles fall out of the transporting medium and settle on a surface. This process is called ***deposition***. **Figure 11g-1** illustrates an area of Death Valley, California where the effects of erosion and deposition can be easily seen.



{PRIVATE**}Figure 11g-1:** The following image was created from DEMs (Digital Elevation Model) for the following 1:24,000 scale topographic quadrangles: Telescope Peak, Hanaupah Canyon, and Badwater, California. To the left is the Panamint Mountain Range. To the right is Death Valley. Elevation spans from 3368 to -83meters and generally decreases from left to right. The *blue line* represents an elevation of 0 meters. Large ***alluvial fans*** extending from a number of mountain valleys to the floor of Death Valley can be seen in the right side of the image. The sediments that make up these ***depositional*** features came from the ***weathering*** and ***erosion*** of bedrock in the mountains located on the left side of the image. (This image was created with ***MacDEM*** software).

Energy of Erosion

The ***energy*** for ***erosion*** comes from several sources. Mountain building creates a disequilibrium within the Earth's landscape because of the creation of relief. ***Gravity*** acts to vertically move materials of higher relief to lower elevations to produce an ***equilibrium***. Gravity also acts on the mediums of erosion to cause them to flow to base level.

Solar radiation and its influence on atmospheric processes is another source of energy for erosion. Rainwater has a ***kinetic energy*** imparted to it when it falls from the ***atmosphere***. Snow has ***potential energy*** when it is deposited in higher elevations. This potential energy can be converted into the energy of motion when the snow is converted into flowing glacial ice. Likewise, the motion of air because of differences in atmospheric pressure can erode surface material when velocities are high enough to cause particle ***entrainment***.

The Erosion Sequence

Erosion can be seen as a sequence of three events: ***detachment***, ***entrainment***, and ***transport***. These three processes are often closely related and sometimes not easy distinguished between each other. A single particle may undergo detachment, entrainment, and transport many times.

Detachment

Erosion begins with the ***detachment*** of a particle from surrounding material. Sometimes detachment requires the breaking of bonds which hold particles together. Many different types of bonds exist each with different levels of particle cohesion. Some of the strongest bonds exist between the particles found within ***igneous rocks***. In these materials, bonds are derived from the growth of ***mineral*** crystals during cooling. In sedimentary rocks, bonds are weaker and are mainly caused by the cementing effect of compounds such as iron oxides, silica, or calcium. The particles found in soils are held together by even weaker bonds which result from the cohesion effects of water and the electro-chemical bonds found in ***clay*** and particles of ***organic matter***.

Physical, ***chemical***, and ***biological weathering*** act to weaken the particle bonds found in rock materials. As a result, weathered materials are normally more susceptible than unaltered rock to the forces of detachment. The agents of erosion can also exert their own forces of detachment upon the surface rocks and soil through the following mechanisms:

Plucking: ice freezes onto the surface, particularly in cracks and crevices, and pulls fragments out from the surface of the rock.

Cavitation: intense erosion due to the surface collapse of air bubbles found in rapid flows of water. In the implosion of the bubble, a micro-jet of water is created that travels with high speeds and great pressure producing extreme stress on a very small area of a surface. Cavitation only occurs when water has a very high velocity, and therefore its effects in nature are limited to phenomenon like high waterfalls.

Raindrop impact: the force of a raindrop falling onto a soil or weathered rock surface is often sufficient to break weaker particle bonds. The amount of force exerted by a raindrop is a function of the ***terminal velocity*** and ***mass*** of the raindrop.

Abrasion: the excavation of surface particles by material carried by the erosion agent. The effectiveness of this process is related to the velocity of the moving particles, their mass, and their concentration at the eroding surface. Abrasion is very active in glaciers where the particles are firmly held by ice. Abrasion can also occur from the particles held in the erosional mediums of wind and water.

Entrainment

Entrainment is the process of particle lifting by the agent of erosion. In many circumstances, it is hard to distinguish between entrainment and detachment. There are several forces that provide particles with a resistance to this process. The most important force is frictional resistance. Frictional resistance develops from the interaction between the particle to its surroundings. A number of factors increase frictional resistance, including: ***gravity***, particle slope angle relative to the flow direction of eroding medium, particle ***mass***, and surface roughness.

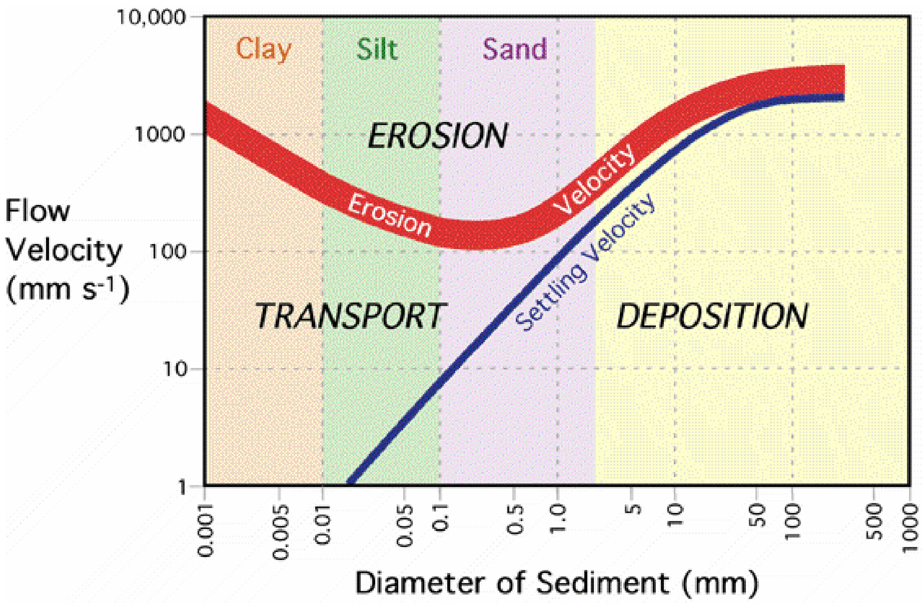
Entrainment also has to overcome the resistance that occurs because of particle cohesive bonds. These bonds are weakened by weathering or forces created by the erosion agent (***abrasion***, ***plucking***, ***raindrop impact***, and ***cavitation***).

Entrainment Forces

The main force reponsible for entrainment is ***fluid drag***. The strength of fluid drag varies with the mass of the eroding medium (water is 9000 times more dense than air) and its velocity. Fluid drag causes the particle to move because of horizontal force and vertical lift. Within a medium of erosion, both of these forces are controlled by velocity. Horizontal force occurs from the push of the agent against the particle. If this push is sufficient to overcome friction and the resistance of cohesive bonds, the particle moves horizontally. The vertical lift is produced by turbulence or eddies within the flow that push the particle upward. Once the particle is lifted the only force resisting its transport is gravity as the forces of friction, slope angle, and cohesion are now non-existent. The particle can also be transported at velocities lower than the entrainment velocities because of the reduction in forces acting on it.

Many hydrologists and geomorphologists require a mathematical model to predict levels of entrainment, especially in stream environments. In these highly generalized models, the level of particle entrainment is relative to particle size and the velocity of the medium of erosion. These quantitative models can be represented graphically. On these graphs, the x-axis represents the log of particle diameter, and the y-axis the log of velocity. The relationship between these two variables to the entrainment of particles is described by a curve, and not by a straight line.

The ***critical entrainment velocity*** curve suggests that particles below a certain size are just as resistant to entrainment as particles with larger sizes and masses (**Figure 11h-2**). Fine silt and clay particles tend to have higher resistance to entrainment because of the strong cohesive bonds between particles. These forces are far stronger than the forces of friction and gravity.



{PRIVATE}Figure 11g-2: This graph describes the relationship between stream flow velocity and particle erosion, transport, and deposition. The curved line labeled "erosion velocity" describes the velocity required to entrain particles from the stream's bed and banks. The erosion velocity curve is drawn as a thick line because the erosion particles tends to be influenced by a variety of factors that changes from stream to stream. Also, note that the entrainment of silt and clay needs greater velocities then larger sand particles. This situation occurs because silt and clay have the ability to form cohesive bounds between particles. Because of the bonding, greater flow velocities are required to break the bonds and move these particles. The graph also indicates that the transport of particles requires lower flow velocities then erosion. This is especially true of silt and clay particles. Finally, the line labeled "settling velocity" shows at what velocity certain sized particles fall out of transport and are deposited.

Transport

Once a particle is entrained, it tends to move as long as the velocity of the medium is high enough to transport the particle horizontally. Within the medium, ***transport*** can occur in four different ways:

Suspension is where the particles are carried by the medium without touching the surface of their origin. This can occur in air, water, and ice.

Saltation is where the particle moves from the surface to the medium in quick continuous repeated cycles. The action of returning to the surface usually has enough force to cause the entrainment of new particles. This process is only active in air and water.

Traction is the movement of particles by rolling, sliding, and shuffling along the eroded surface. This occurs in all erosional mediums.

Solution is a transport mechanism that occurs only in aqueous environments. Solution involves the eroded material being dissolve and carried along in water as individual ions.

Particle weight, size, shape, surface configuration, and medium type are the main factors that determine which of these processes operate.

Deposition

The ***erosional*** transport of material through the landscape is rarely continuous. Instead, we find that particles may undergo repeated cycles of ***entrainment***, ***transport***, and ***deposition***. Transport depends on an appropriate balance of forces within the transporting medium. A reduction in the velocity of the medium, or an increase in the resistance of the particles may upset this balance and cause deposition. Reductions in competence can occur in a variety of ways. Velocity can be reduced locally by the sheltering effect of large rocks, hills, stands of vegetation or other obstructions. Normally, competence changes occur because of large scale reductions in the velocity of flowing medium. For wind, reductions in velocity can be related to variations in spatial heating and cooling which create pressure gradients and wind. In water, lower velocities can be caused by reductions in discharge or a change in the grade of the stream. Glacial flows of ice can become slower if precipitation input is reduced or when the ice encounters melting. Deposition can also be caused by particle ***precipitation*** and ***flocculation***. Both of these processes are active only in water. Precipitation is a process where dissolved ions become solid because of changes in the temperature or chemistry of the water. Flocculation is a chemical process where salt causes the aggregation of minute clay particles into larger masses that are too heavy to remain suspended.

(h) Hillslope Processes and Mass Movement

{PRIVATE}Introduction

Hillslopes are an important part of the terrestrial landscape. The Earth's landscape can be thought of as being composed of a mosaic of slope types, ranging from steep mountains and cliffs to almost flat plains. On most hillslopes large quantities of ***soil*** and ***sediment*** are moved over time via the mediums of air, water, and ice often under the direct influence of ***gravity***. The form a hillslope takes is dependent on the various geomorphic processes acting on it. Hillslopes are also the source of materials that are used to construct a number of ***depositional landforms***.

In practical terms, hillslopes have direct and indirect influence on a number of human activities. The steepness and structural stability of hillslopes determines their suitability for agriculture, forestry, and human settlement. Hillslopes can also become a ***hazard*** to humans if their materials move rapidly through the process of ***mass wasting***.

Inputs and Outputs to the Hillslope System

We can begin our study of hillslopes by thinking of them as a ***process-response system***. The hillslope system receives inputs of ***solar radiation***, ***precipitation***, solid and dissolved substances from the ***atmosphere***, and unconsolidated sediment derived from the ***weathering*** of bedrock. The inputs of unconsolidated sediment are controlled by weathering rates. In general, the warmer the climate the higher the rates of bedrock weathering. Rates of weathering are also influenced by the presence of moisture.

Outputs to hillslopes occur by ***evapotranspiration***, by ***percolation*** of water and the movement of dissolved substances into the bedrock, and by removal of ***sediment*** by ***streams*** by ***glaciers*** or by ***ocean*** waves and currents. Outputs of debris or sediment from hillslope systems are controlled primarily by the availability of ***erosional*** mechanisms to transport material that accumulates at the slope's surface and base. For example, the presence of a stream at the base of a hillslope encourages removal of sediment that moves downslope. If the stream's ***discharge*** is too small to handle the debris input, sediment will accumulate at the base of the slope.

The magnitude of hillslope inputs and outputs depends upon a number of factors, including bedrock geology, climate, and the nature of the slope to the broader landscape. The balance between inputs and outputs from the hillslope system exerts a major control over the form of the developing slope. In situations where inputs are the controlling factor, the slope is said to be weathering limited because outputs quickly remove any accumulating debris. Where the potential for weathering is high but outputs are restrained the hillslope system is classified as being transport limited. Landscapes that are transport limited are easily recognized by the presence of a deep ***soil profile***.

Mass Movement and Hillslope Stability

A variety of processes exist by which materials can be moved through the hillslope system. These processes are generically known as ***mass movement*** or ***mass wasting***. The operation of mass movement processes relies upon the development of instability in the hillslope system. Under these conditions, failure of the slope material can occur on a range of time scales. Some types of mass movement involve rather rapid, spontaneous events. Sudden failures tend to occur when the stresses exerted on the slope materials greatly exceed their strength for short periods of time. In many cases, type of mass movement is produced by the operation of short term trigger mechanisms. Mass movement can also be a less continuous process that occurs over long periods of time. Slow failures often occur when the applied stresses only just exceed the internal strength of the hillslope system.

What are the sources of the stresses and strength acting within hillslope materials? As we have noted, a major source of stress is the gravitational force. The magnitude of this force is related to the angle of the slope and the weight of hillslope sediments and rock. The following equation models this relationship:

$F = W \sin Ø$

where **F** is gravitational force, **W** is the weight of the material occurring at some point on the slope, and **Ø** is the angle of the slope.

The internal strength of the hillslope system varies according to the nature of the materials making up the slope. Hillslopes composed of loose materials, like sand and gravel, derived their internal from frictional resistance which depends on the size, shape, and arrangement of the particles. Hillslopes consisting of ***silt*** and ***clay*** particles obtain their internal strength from particle cohesion which is controlled by the availability of moisture in the soil. Too much moisture breaks the cohesive bonds and can turn a solid hillslope into a river of mud. ***Rock*** slopes generally have the greatest internal strength. Internal strength in these systems is derived from the effects of the solidification and crystallization of ***magma*** or the ***lithification*** of once loose particles.

The stability of a hillslope depends on the relationship between the stresses applied to the materials that make up the slope and their internal strength. Mass movement occurs when the stresses exceed the internal strength. This condition is not always caused by an increase in stress. In some cases, the internal strength of the materials can be reduced over relatively short periods of time resulting in mass movement.

Many factors can act as triggers for hillslope failure. One of the most common is prolonged or heavy rainfall. Rainfall can lead to mass movement through three different mechanisms. Often these mechanisms do not act alone. The saturation of soil materials increases the weight of slope materials which then leads to greater gravitational force. Saturation of soil materials can reduce the cohesive bonds between individual soil particles resulting in the reduction of the internal strength of the hillslope. Lastly, the presence of bedding planes in the hillslope material can cause material above a particular plane below ground level to slide along a surface lubricated by ***percolating*** moisture.

Earthquakes are another common mechanism that can trigger mass movement. The ***seismic waves*** produced by earthquakes vibrate slope materials. This vibration can lead to failure by increasing the downward stress or by decreasing the internal strength of the hillslope sediments through particle movement.

Water, Sediment Transport, and Hillslopes

Rainsplash is a microscale process that can be quite effective in moving material on slopes. The impact of ***rain*** droplets on the ***soil*** surface often detaches individual grains of soil moving them some distance from their source. On flat surfaces, the effect of rain drop impact is to redistribute the material without any net transport in a particular direction. However, on a slope the influence of gravity and slope encourage more material to be redistributed downslope rather than upslope. When slopes become 25 degrees or greater, almost all the redistribution occurs in a downslope direction.

Considerable transport of surface sediments on slopes occurs by ***rainwash*** and surface ***runoff***. On relatively flat surfaces, runoff occurs as a continuous layer of water commonly called ***sheetwash***. The erosive potential of sheetwash is usually quite limited because this type of flow is shallow and non-turbulent and cannot readily

entrain surface particles. However, topographic irregularities can quickly transform sheetwash into small channels called ***rills***. Rills then coalesce into larger ***stream channels*** and so on. Rills and large stream channels concentrate the movement of water causing an increase in flow velocity and turbulence. Higher flow velocities and turbulence lead to a greater potential for entrainment and subsequent transport of hillslope materials.

Mass Movement in Non-Cohesive Materials

Many slopes are composed of non-cohesive, coarse-grained sediments. This type of slope is common to landform features like ***alluvial fans***, ***screes***, ***talus cones***, ***sand dunes***, and ***glacial outwash*** deposits. On slopes of this type, mass movement often occurs through the sliding or rolling of a small number of particles as localized instabilities develop (**Figure 11h-1**). In some cases, these movements can organize themselves into larger avalanches through a domino effect. Mass movement on non-cohesive materials can also occur by way of shallow sliding. Shallow sliding occurs when planes of weakness develop just beneath the surface of the slope. Planes of weakness develop where horizontal layering occurs in the sediment. This layering can be caused by the nature of sediment deposition, percolation of water, or by the presence of subsurface soil, sediment or rock layers.

{PRIVATE**}Figure 11h-1:** ***Scree*** slope formed of non-cohesive sediments at the base of a steep hill. Mass movement on this type of hillslope occurs primarily by way of localized sliding or rolling of a relatively small number of particles.

In many mountainous areas, shallow coarse soils develop over bedrock. Under the right consitions, large downslope movements of this material can occur. This type of mass movement is known as a ***debris flow***.

Mass Movement in Cohesive Materials

Slopes formed from ***clays*** and ***silts*** sediments display somewhat unique ***mass movement*** processes. Clay and silt particles have a degree of cohesion which gives them potentially more internal strength than non-cohesive sediments. This cohesion occurs because of electrochemical bonds which operate between particles and the surface tension effects of water films that forms around particles. Both of these sources of cohesion are dependent upon moisture content. Maximum cohesiveness takes place when mositure conditions are moderate. Too much or too little water reduces the strength of the cohesion.

Two common types of mass movements in cohesive materials are ***rotational slips*** and ***mudflows***. Both of these processes occur over very short time periods. Rotational slips or slumps occur along clearly defined planes of weakness which generally have a concave form beneath the Earth's surface (**Figure 11h-2**). Rotational slips can be caused by a variety of factors. The most common mechanism reason for them to occur is erosion at the base of the slope which reduces the support for overlying sediments. Erosion at the base of a slope can be caused by the presence of a stream channel or by wave action.

{PRIVATE**}Figure 11h-2:** The head of a ***rotational slip*** in the Black Hills of North Dakota. (**Source:** Image provided by the *National Geophysical Data Center*, NOAA).

Mudflows occur when slope materials become so saturated that the cohesive bonds between particles is lost. The saturated material then flows like a thick fluid downslope (**Figure 11h-3**). Flow stops when water loss through seepage causes the sediment to solidify. Mudflows can occur on very low slope angles because internal particle frictional resistance and cohesion is negligible.

{PRIVATE**}Figure 11h-3:** The Slumgullian ***mudflow*** in the San Juan Mountains, Colorado. (**Source:** Image provided by the *National Geophysical Data Center*, NOAA).

Some of the mass movement processes operating on cohesive materials occur over very long time spans. One of the most widespread of these processes is ***soil creep***. Soil creep involves the movement of slope sediments in a series of numerous cyclical steps (**Figure 11h-4**). The cyclical effects of temperature fluctuations, variations in moisture, and gravity on inclined soil sediments often cause this process.

Solifluction is the slow movement of soil caused by ***freeze-thaw action***. This process is a widespread in polar and sub-polar regions where exists. Solifluction occurs when seasonal or daily fluctuations of temperature are above freezing. At these temperatures, the upper portion of the soil surface and permafrost thaw creating waterlogged mass because subsurface ice prevents drainage. The waterlogged mass then flows downslope as lobes of sediment and surface vegetation.

Mass Movement on Hard Rock Slopes

Mass movement on hard rock slopes is often dramatic and quick (**Figure 11h-5**). Hard rocks derive their internal strength mainly from the strong inter-granular bonds that form when ***magma*** cools and crystallizes or when ***lithification*** occurs in ***sedimentary rocks***. Because of their strong internal strength hard rock slopes can have relatively steep angles. Nevertheless, weaknesses do occur along ***bedding*** planes and ***joints*** naturally found in these materials. Most mass movement on hard rock slopes involves the downward movement of small rock fragments pried loose by gravitational stress and/or ***freeze-thaw processes***. We call these types of mass wasting ***rockfalls***.

{PRIVATE**}Figure 11h-5:** ***Rockfalls*** from steep hillsides frequently disrupt road traffic in the mountainous regions of British Columbia, Canada. (**Source:** Image provided by Natural Resources Canada - *Geoscape Vancouver*).

Larger scale downslope movement of rock can also occur along well defined ***joints*** or ***bedding*** planes. This type of movement is called a ***rock slide***. Most rock slides occur because the hillslope and rock layers ***dip*** in the same direction (**Figure 11h-6**). In these cases, the rock slide occurs when a fracture plane develops causing overlying materials to slide downslope.

{PRIVATE**}Figure 11h-6:** ***Rock slide*** above Summit Lake, British Columbia, Canada. This rock slide occurred on April 2, 1999 at about 10:00 o'clock in the evening and involved the movement of about 200,000 cubic meters of material.

Not all rock slides are the result of the process described above. The rock slide that covered the town of Frank, Alberta, Canada on April 29, 1903 was in part caused by human activities (**Figure 11h-7**). The sedimentary rock layers in Turtle Mountain ***dipped*** away from the valley containing the mining town of Frank. However, ***joints*** in the limestone layers did dip toward the town. The mountain was also composed of some structurally weak limestone, shale, siltstone, and coal layers that were deformed by the weight of more massive limestone located above. Finally, the mining of coal at the Turtle Mountain's base reduced the support to overlying materials. Together these factors lead to the sudden movement of 33 million cubic meters of rock in approximately two minutes.

{PRIVATE**}Figure 11h-7:** The ***rock slide*** at Frank, Alberta, Canada (1903) moved 33 million cubic meters of rock from Turtle Mountain over the town of Frank in less than two minutes killing 70 people. (**Source:** Image provided by the *National Geophysical Data Center*, NOAA).

(i) Streamflow and Fluvial Processes

Streams alter the Earth's landscape through the movement of water and ***sediment*** (**Figure 11i-1**). Streams are powerful ***erosive*** agents moving material from their ***bed*** and ***banks***. In mountainous regions, stream erosion often produces deep ***channels*** and ***canyons***. Streams also deposit vast amounts of sediment on the terrestrial landscape and within ***lakes*** and ***ocean basins***.

{PRIVATE}**Figure 11i-1:** Portion of stream channel located in the headwaters. Steep elevation gradient makes this part of the stream very effective at erosion and transportation of sediment.

Geomorphologists often view streams as ***systems***. The stream system, like almost all environmental systems, is open to both inputs and outputs of various types of materials. Water enters the stream system by direct ***precipitation*** in the channel, from ***runoff***, ***throughflow***, and by ***groundwater flow***. The movement of water into a stream also carries with it dissolved and solid materials eroded from the surrounding landscape, stream banks, and the stream bed. Sediments carried by streams to lower elevations are occasionally deposited and stored at numerous locations for various periods of time within the stream system before they reach their final resting place.

Losses of material carried by the stream system occurs through a number of processes. Water is lost by ***evaporation***, ***seepage***, and ***flooding***. Streamflow ends when the water carried by stream enters a receiving basin like a ***lake*** or an ***ocean***. Sediment is lost by various types of ***deposition***.

The Long Profile of Streams

The topographic ***long profile*** or grade of an average ***stream*** is concave-upwards (**Figure 11i-2**). At their ***headwaters***, the grade of a stream is usually steep. As streams get closer to sea-level, the angle of the grade becomes more gently sloping. Near the ***mouth*** of the stream, the grade becomes almost flat.

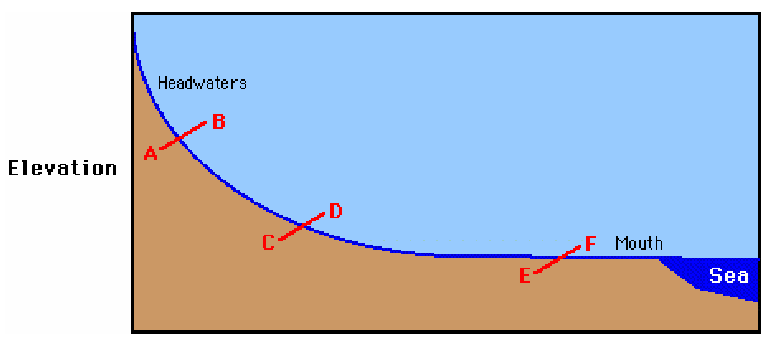


Figure 11i-2: Long profile of a typical stream.

The grade of a stream develops over thousands and sometimes millions of years. It is an equilibrium process that attempts to reduce topographic bumps in the long profile through ***erosion*** and ***deposition***. This process is without end as crustal uplift, due to ***plate tectonics***, is always modifying the topographic nature of the Earth's landscape.

The following **diagrams** and **photographs** illustrate the nature of the stream channel at three different locations along the stream profile. The first illustration describes the channel at point **A-B** (**Figure 11i-3**). This channel is located in the stream's headwaters. The stream gradient and surrounding topography is quite steep. Stream velocity is at a maximum and cuts a narrow deep channel. The ***floodplain*** is minimal as high velocities carry all sizes of load down stream. ***Stream load*** is high as the extreme relief provides the energy for extensive ***erosion*** producing a V-shaped river valley.

{PRIVATE**}**

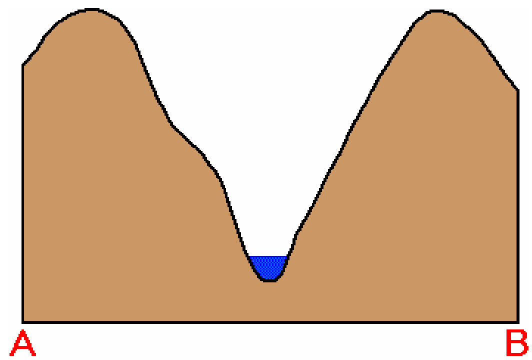


Figure 11i-3: Stream channel near the headwaters.

Further down the profile, changes in relief causes the nature of the ***stream channel*** to modify (**Figure 11i-4:** diagram **C-D**). The channel is now leaving the mountainous terrain and entering a landscape that is more gently sloping. This change causes a sudden reduction in the stream's velocity. The stream adjusts to this change, by depositing most of its coarse ***stream load*** onto the ***floodplain***. The stream also takes on a ***braided*** channel form. These channels are always changing in size, number, and location because of temporal variations in ***stream discharge***. The amount of sediment in the floodplain also varies significantly over time because of temporal fluctuations in flow.

{PRIVATE**}**

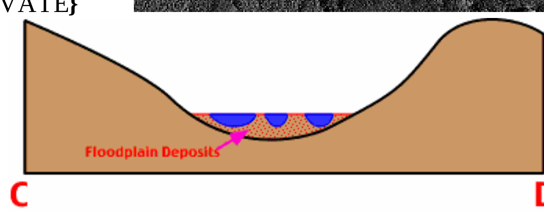
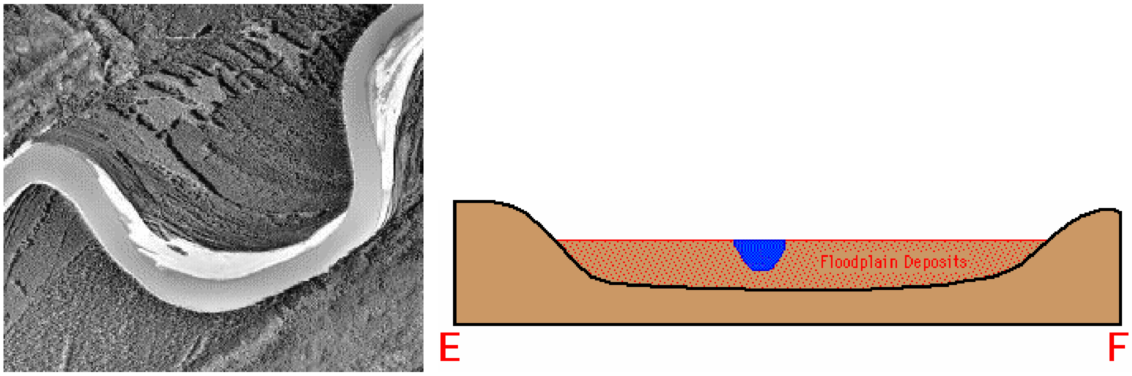


Figure 11i-4: Stream channel near the middle of a typical stream profile.

The final image and illustration describes the channel near the ***mouth*** of the stream (**Figure 11i-5**: diagram **E-F**). The extensive flat floodplain is composed of ***flooding*** deposits or ***point bar*** deposits from channel ***meandering***. The channel is quite large and U-shaped. ***Stream discharge*** is at a maximum and sediment load is generally composed of finer materials.



{PRIVATE**}**

Figure 11i-5: Stream channel near the mouth of a stream.

Stream Discharge

The flow of water through a ***stream channel*** is called ***stream discharge***. In most countries, it is measured in cubic meters per second. The following equation defines stream discharge mathematically:

$Q = V \times W \times D$

where **Q** is the discharge, **V** is the velocity, **W** is the average width and **D** is the average depth of the flow. Stream discharge varies over both time and space. Discharge normally increases downstream as more water enters the stream channel from ***runoff*** and ***groundwater flow***. Discharge varies temporally because of chaotic behavior of its inputs like ***precipitation*** and ***snow melt***. As discharge increases corresponding changes in velocity, channel depth and width are made within the stream system. Of the three variables that change within the stream system with an increase in discharge, velocity is the least responsive.

Velocity and Turbulence

Because of ***frictional*** drag, stream velocity is at a maximum at the center of the channel near the surface and a minimum near the ***bed*** and ***banks***. If we examine stream velocity in three dimensions, a more complex pattern becomes apparent. The line of maximum velocity or ***thalweg*** moves from side to side within a channel.

The dynamics of stream flow is primarily influenced by ***friction***, channel topography and channel shape. Within a stream channel, three types flow can be observed:

Laminar flow - water flow in the stream is not altered in its direction. Water flows as parallel molecular streams.

Turbulent flow* -** water flows as discrete ***eddies and ***vortices***. Caused by channel topography and ***friction***.

Helical flow* -** spiral flow in a stream. Caused by channel shape. ***Meandering channels cause this type of flow. Helical flow has an important role in ***sediment*** transport and ***deposition***, and in the creation of ***point bars***.

Finally, flow is not always contained within the stream channel. During periods of high stream discharge ***overbank flow*** may occur. Overbank flow or flooding involves the spilling of water over the ***stream's banks*** and onto the ***floodplain***.

Sediment Transport

All streams carry ***sediment***. Most of the sediment found in a stream has been washed into the channel from surface ***runoff***. Sediment is also added from the ***erosion*** of the stream channel ***bed*** and ***banks***. The quantity of sediment in a stream varies temporally due to changes in discharge. Normally, as ***discharge*** and velocity increase, the amount of sediment being carried by the stream rises correspondingly.

For many discharge stations, scientists have determined the quantitative relationship between stream discharge and concentration of sediment transported by the stream. This information is often displayed graphically as a

sediment rating curve. The slope of such curves tend to vary with each stream as each stream system has its own unique environmental characteristics.

During periods of low discharge, very little sediment movement takes place, and material found in the beds and banks of river channels tend to be stable. When discharge increases, more and more sediment is eroded from the stream bed and stream banks. Loose sediment on the bed is picked up as the ***fluid drag*** of the flowing water increases. High levels of discharge also causes significant amounts of material to be added to the flow of the stream from the erosion of the stream bank. This process is called ***bank-caving*** by hydrologists.

Streams generally transport three types of material: ***bed load*** (pebbles and ***sand*** which move along the stream bed without being permanently suspended in the flowing water), ***suspended load*** (***silts*** and ***clays*** in suspension) and ***dissolved load*** (material in ***solution***) (**Figure 11i-6**). The absolute quantities and the relative proportions of these types of ***stream load*** vary from one stream to another, and within a single stream from one time to another.

Along the stream bed particles are moved via ***traction*** (sliding and rolling) and ***saltation***. Saltation is a process where particles moves from the stream bed to the medium in quick continuous repeated cycles. The action of returning to the surface usually has enough force to cause the ***entrainment*** of new particles. Particles that are transported as ***suspended load*** remain entrained in the flowing water for long periods of time. These particles can be deposited to the stream bed when stream velocity is reduced Human activity adds large amounts of dissolved and solid material to streams. These added materials include fertilizers, animal waste, and soluble compounds that are the by products of agriculture, forestry, and industry. Agriculture and forestry also add large amounts of solid sediment into streams because of their ***disturbance*** of vegetated surfaces.

(j) Fluvial Landforms

{PRIVATE}Stream Channel Types

Within a single ***stream*** we can often recognize three different ***channel*** types. These unique channel types develop in response to changes in stream velocity, ***sediment*** texture, and stream grade.

Channels located in the upper reaches of many streams tend to be narrow with flow moving at high velocities (**Figure 11j-1**). The high flow velocities found in these streams are the result of a steep grade and gravity. Within these stream systems, ***erosion*** is a very active process as the channel tries to adjust itself to the topography of the landscape. ***Deposition*** occurs primarily during periods of low flow. As a result, ***floodplain*** deposits are very limited, and the ***stream bed*** is very transient and shallow.

Figure 11j-1: Upper reach of a stream in the Rocky Mountains, Canada.

Streams with high sediment loads that encounter a sudden reduction in flow velocity generally have a ***braided*** channel type (**Figure 11j-2**). This type of stream channel often occurs further down the ***stream profile*** where the grade changes from being steep to gently sloping. In a braided stream, the main channel divides into a number of smaller, interlocking or braided channels. Braided channels tend to be wide and shallow because bedload materials are often coarse (***sands*** and gravels) and non-cohesive.

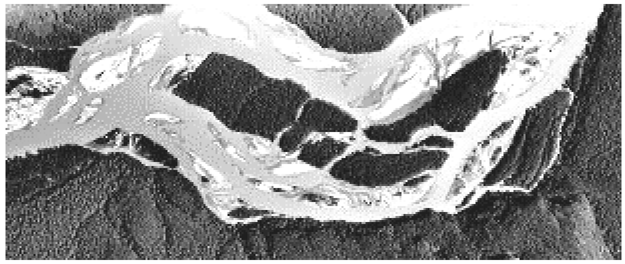


Figure 11j-2: Braided stream channel.

Meandering channels form where streams are flowing over a relatively flat landscape with a broad floodplain (**Figure 11j-3**). Technically, a stream is said to be meandering when the ratio of actual channel length to the straight line distance between two points on the stream channel is greater than 1.5. Channels in these streams are characteristically U-shaped and actively migrate over the extensive ***floodplain***.

Figure 11j-3: Meandering stream channel.

Stream Channel Features

Within the stream channel are a variety of sedimentary ***beds*** and structures. Many of these features are dependent upon the complex interaction between stream velocity and ***sediment*** size.

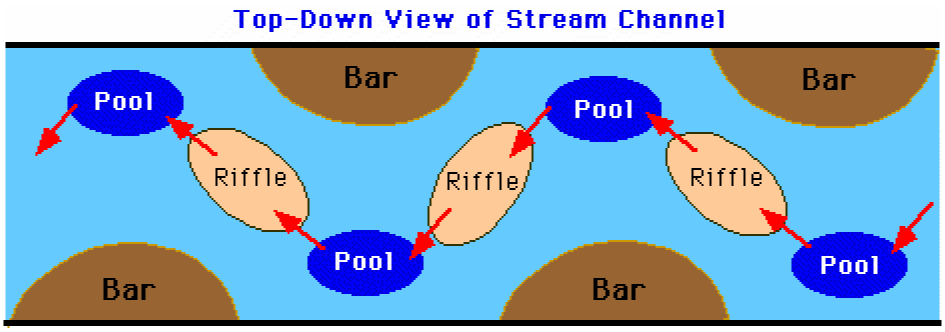
Streams carrying coarse sediments develop sand and gravel ***bars***. These types of bars seen often in ***braided streams*** which are common in elevated areas (**Figure 11j-4**). Bars develop in braided streams because of reductions in ***discharge***. Two conditions often cause the reduction in discharge: reduction in the gradient of the stream and/or the reduction of flow after a precipitation event or spring melting of snow and ice.

Figure 11j-4: Braided stream channel with gravel bars.

Point bars develop where stream flow is locally reduced because of friction and reduced water depth (**Figure 11j-5**). In a ***meandering*** stream, point bars tend to be common on the inside of a channel bend.

Figure 11j-5: Meandering stream channel as seen from above.

In straight streams, bar-like deposits can form in response to the ***thalweg*** (**red** arrows **Figure 11j-6**) and ***helical flow***. **Figure 11j-6** below shows an overhead view of these deposits and related features.



{PRIVATE}**Figure 11j-6:** Overhead view of the depositional features found in a typical straight stream channel.

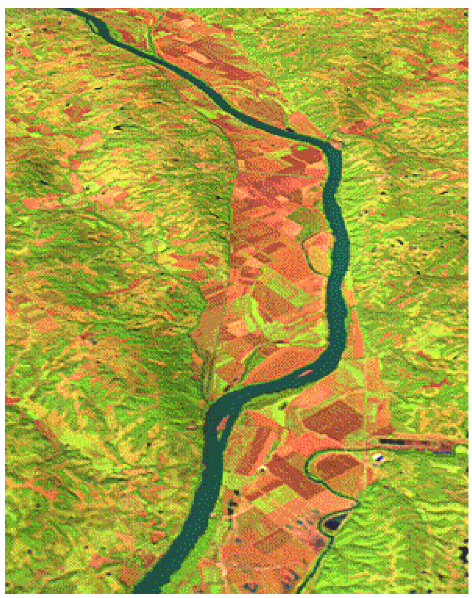
In this straight channel stream, ***bars*** form in the regions of the stream away from the thalweg. ***Riffles***, another type of coarse deposit, develop beneath the thalweg in locations where the faster flow moves vertically up in the channel. Between the riffles are scoured ***pools*** where material is excavated when the zone of maximum stream velocity approaches the stream's bed. The absolute spacing of these features varies with the size of the channel.

However, the relative distance between one riffle and the next is on average five to seven times the width of the channel (exaggerated in diagram). Both of these features can also occur in sinuous channels.

Dunes and ***ripples*** are the primary sedimentary features in streams whose channel is composed mainly of sand and silt. Dunes are about 10 or more centimeters in height and are spaced a meter or more apart. They are common in streams with higher velocities. Ripples are only a few centimeters in height and spacing, and are found in slow moving streams with fine textured beds. Both of these features move over time, migrating down stream. Material on the gently sloping ***stoss-side*** of these features rolls and jumps up the slope under the influence of water flow. Particles move up the slope until they reach the crest of the feature and then avalanche down the steeper ***lee-side*** to collect at the base of the next dune or ripple. This process is then repeated over and over again until the material reaches a location down stream where it is more permanently deposited.

The Floodplain

Alongside ***stream channels*** are relatively flat areas known as ***floodplains*** (**Figure 11j-7**). Floodplains develop when streams over-top their ***levees*** spreading ***discharge*** and suspended ***sediments*** over the land surface during ***floods***. Levees are ridges found along the sides of the stream channel composed of sand or gravel. Levees are approximately one half to four times the channel width in diameter. Upon retreat of the flood waters, stream velocities are reduced causing the deposition of ***alluvium***. Repeated flood cycles over time can result in the deposition of many successive layers of alluvial material. Floodplain deposits can raise the elevation of the ***stream bed***. This process is called ***aggradation***.



{PRIVATE**}Figure 11j-7:** The following Landsat 5 image taken in September 1992 shows a section of the Missouri River at Rocheport, Missouri. The oblique perspective of this image is looking westward or upstream. This image has been color enhanced and modified to show an exaggerated topographic relief. Bare soil and plowed land appears red, vegetation appears green, and water is dark blue. A flat river flood plain can be seen in the center of the image. Because of the season, most of the farmland located on the rich and fertile soils of the floodplain is plowed and devoid of vegetation. (**Source:** *NASA Scientific Visualization Studio*).

Floodplains can also contain sediments deposited from the lateral migration of the river channel. This process is common in both braided and meandering channels. ***Braided*** channels produce horizontal deposits of sand during times of reduced ***discharge***. In ***meandering*** streams, channel migration leads to the vertical deposition of ***point bar*** deposits. Both braided and meandering channel deposits are more coarse than the materials laid down by flooding.

A number of other geomorphic features can be found on the floodplain. Intersecting the levees are narrow gaps called ***crevasses*.** These features allow for the movement of water to the floodplain and back during floods. Topographical ***depressions*** are found scattered about the floodplain. Depressions contain the some of the finest deposits on the floodplain because of their elevation. ***Oxbow lakes*** are the abandoned channels created when meanders are cut off from the rest of the channel because of lateral stream ***erosion***.

Alluvial Fans and Deltas

Streams flowing into standing water normally create a ***delta*** (**Figure 11j-8** and **11j-9**). A delta is body of ***sediment*** that contains numerous horizontal and vertical layers. Deltas are created when the sediment load carried by a stream is deposited because of a sudden reduction in stream velocity. The surface of most deltas is marked by small shifting channels that carry water and sediments away from the main river channel. These small channels also act to distribute the stream's sediment load over the surface of the delta. Some deltas, like the Nile, have a triangular shape. Streams, like the Mississippi, that have a high sediment content and empty into relatively calm waters cause the formation of a birdfoot shaped delta**.**

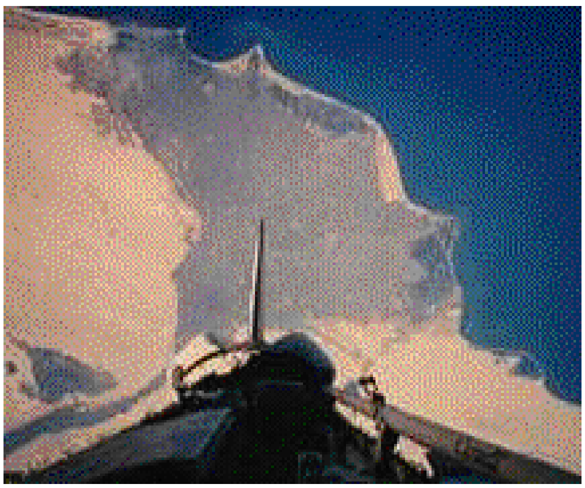


Figure 11j-8: Nile Delta (**Source:** *NASA*).

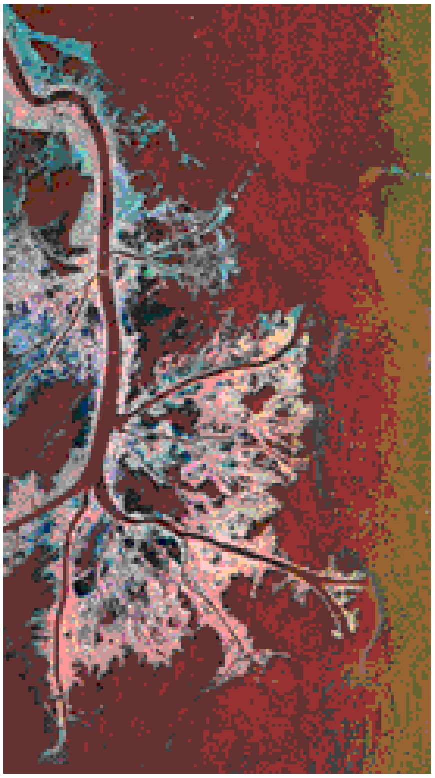
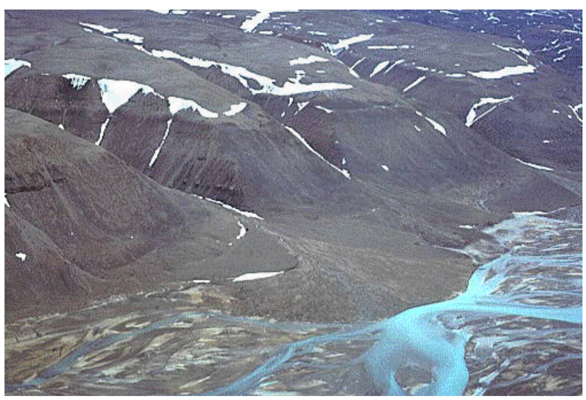


Figure 11j-9: Mississippi Birdfoot Delta (**Source:** *NASA*).

Most deltas contain three different types of deposits: ***foreset***, ***topset*** and ***bottomset*** beds. ***Foreset beds*** make up the main body of deltas. They are deposited at the outer edge of the delta at an angle of 5 to 25 degrees. Steeper angles develop in finer sediments. On top of the foreset beds are the nearly horizontal ***topset beds***. These beds are of varying grain sizes and are formed from deposits of the small shifting channels found on the delta surface. In front and beneath the foreset beds are the ***bottomset beds***. These beds are composed of fine ***silt*** and ***clay***. Bottom set beds are formed when the finest material is carried out to sea by stream flow.

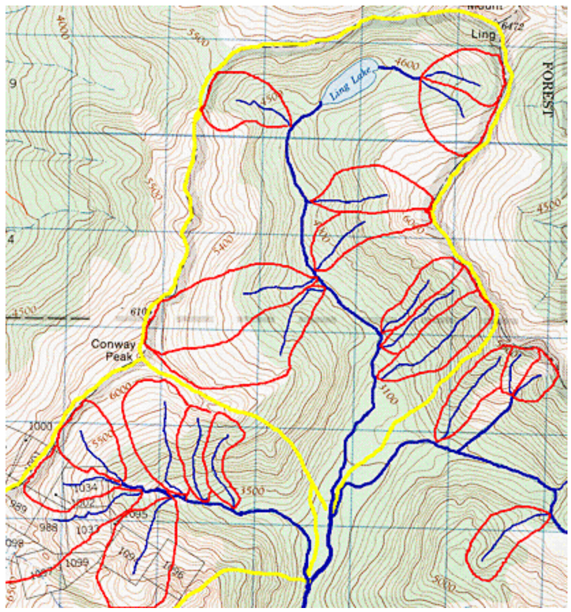
An ***alluvial fan*** is a large fan-shaped deposit of sediment on which a ***braided stream*** flows over (**11j-10**). Alluvial fans develop when streams carrying a heavy load reduce their velocity as they emerge from mountainous terrain to a nearly horizontal plain. The fan is created as braided streams shift across the surface of this feature depositing sediment and adjusting their course. The image below shows several alluvial fans that formed because of a sudden change in elevation.



{PRIVATE**}Figure 11j-10:** Alluvial Fans - Brodeur Peninsula, Baffin Island, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

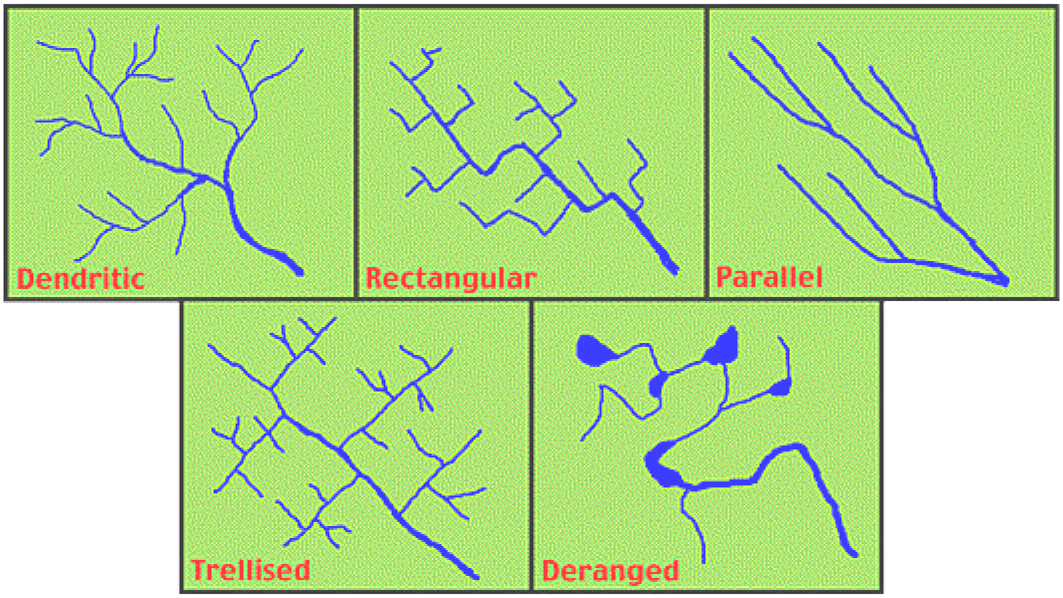
(k) The Drainage Basin Concept

{PRIVATE}Geomorphologists and hydrologists often view ***streams*** as being part of ***drainage basins***. A drainage basin is the topographic region from which a stream receives ***runoff***, ***throughflow***, and ***groundwater flow***. Drainage basins are divided from each other by topographic barriers called a ***watershed*** (**Figure 11k-1**). A watershed represents all of the stream tributaries that flow to some location along the stream channel. The number, size, and shape of the drainage basins found in an area varies with the scale of examination. Drainage basins are arbitrarily defined based on the topographic information available on a map. The quality of this information decreases as map scale becomes smaller (see **topic** *2a*).



{PRIVATE}**Figure 11k-1:** The following image shows the nested nature of drainage basins as determined from a topographic map sheet. The red lines describe the watersheds for the drainage basins of first order streams. The yellow lines define the watersheds for two drainage basins from locations further upstream. Note that the first order basins are components of these much large drainage basins.

Drainage basins are commonly viewed by scientists as being open systems. Inputs to these systems include ***precipitation***, ***snow melt***, and ***sediment***. Drainage basins lose water and sediment through ***evaporation***, ***deposition***, and streamflow. A number of factors influence input, output, and transport of sediment and water in a drainage basin. Such factors include topography, soil type, bedrock type, climate, and vegetation cover. These factors also influence the nature of the pattern of stream channels (**Figure 11k-2**).



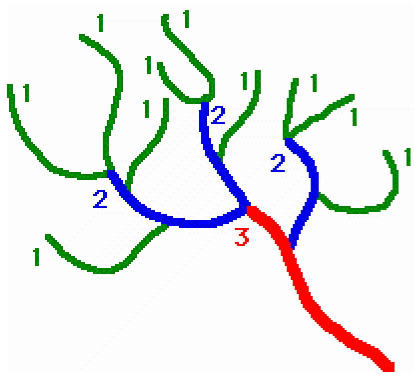
{PRIVATE}**Figure 11k-2:** Common drainage pattern types.

Trellised drainage patterns tend to develop where there is strong structural control upon streams because of geology. In such situations, channels align themselves parallel to structures in the bedrock with minor tributaries coming in at right angles. Areas with tectonic faults or bedrock joints can cause streams to take on a grid-like or **rectangular** pattern. **Parallel** drainage patterns are often found in areas with steep relief or where flow is over non-cohesive materials. **Dendritic** patterns are typical of adjusted systems on erodable sediments and uniformly dipping bedrock. **Deranged** drainage patterns are found in areas recently disturbed by events like glacial activity or volcanic deposition. Over time, the stream will adjust the topography of such regions by transporting sediment to improve flow and channel pattern.

(l) Stream Morphometry

{PRIVATE}***Morphometry*** is defined as the measurement of the shape. Morphometric studies in the field of hydrology were first initiated by R.E. Horton and A.E. Strahler in the 1940s and 1950s. The main purpose of this work was to discover holistic ***stream*** properties from the measurement of various stream attributes.

One of the first attributes to be quantified was the hierarchy of stream segments according to an ordering classification system (**Figure 11l-1**). In this system, channel segments were ordered numerically from a stream's ***headwaters*** to a point somewhere down stream. Numerical ordering begins with the tributaries at the stream's headwaters being assigned the value 1. A stream segment that resulted from the joining of two 1st order segments was given an order of 2. Two 2nd order streams formed a 3rd order stream, and so on. Analysis of this data revealed some interesting relationships. For example, the ratio between the number of stream segments in one order and the next, called the ***bifurcation ratio***, was consistently around three. R.E. Horton called this association the ***law of stream numbers***.



{PRIVATE}Order	Number of Segments	Bifurcation Ratio
1	10	
		{PRIVATE}**3.33**
2	3	**3.00**
3	1	

Figure 11l-1: Example of stream ordering and the calculation of bifurcation ratio.

R.E. Horton applied morphometric analysis to a variety of stream attributes and from these studies he proposed a number of **laws of drainage composition**. Horton's ***law of stream lengths*** suggested that a geometric relationship existed between the number of stream segments in successive stream orders. The ***law of basin areas*** indicated that the mean basin area of successive ordered streams formed a linear relationship when graphed. The results described above and the outcomes of other related analyses convinced researchers that these findings suggested that some underlying factor (or factors) was governing the structure of the various stream attributes in a similar predictable way. Studies of other natural branching networks have revealed patterns similar to the stream order model. For example, the bifurcation ratio of three has also been discovered in the rooting systems of plants, the branching structure of woody plants, and the veination in leaves and the human circulatory system.

In addition to the mathematical relationships found in stream ordering, various aspects of drainage network forms were also found to be quantifiable. One such relationship was ***drainage density***. Drainage density is a measure of the length of ***stream channel*** per unit area of drainage basin. Mathematically it is expressed as:

Drainage Density (Dd) = Stream Length / Basin Area

The measurement of drainage density provides a hydrologist or geomorphologist with a useful numerical measure of landscape dissection and ***runoff*** potential. On a highly permeable landscape, with small potential for runoff, drainage densities are sometimes less than 1 kilometer per square kilometer. On highly dissected surfaces densities of over 500 kilometers per square kilometer are often reported. Closer investigations of the processes responsible for drainage density variation have discovered that a number of factors collectively influence stream density. These factors include climate, topography, soil ***infiltration*** capacity, vegetation, and geology.

(m) Coastal and Marine Processes and Landforms

{PRIVATE}The various landforms of coastal areas are almost exclusively the result of the action of ***ocean waves***. Wave action creates some of the world's most spectacular ***erosional landforms***. Where wave ***energy*** is reduced ***depositional landforms***, like ***beaches***, are created.

Properties of Waves

The source of ***energy*** for coastal ***erosion*** and ***sediment transport*** is ***wave*** action. A wave possesses ***potential energy*** as a result of its position above the ***wave trough***, and ***kinetic energy*** caused by the motion of the water within the wave. This wave energy is generated by the ***frictional*** effect of winds moving over the ocean surface.

The higher the wind speed and the longer the ***fetch***, or distance of open water across which the wind blows and waves travel, the larger the waves and the more energy they therefore possess. It is important to realize that moving waves do not move the water itself forward, but rather the waves impart a circular motion to the individual molecules of water. If you have ever gone fishing in a boat on the ocean or a large lake you will have experienced this phenomenon. As a moving wave passes beneath you, the boat rises and falls but does not move any distance across the water body.

Waves posses several measurable characteristics including length and height. ***Wavelength*** is defined as the horizontal distance from ***wave crest*** to wave crest, while ***wave height*** is the vertical difference between the wave's ***trough*** and ***crest***. The time taken for successive crests to pass a point is called the ***wave period*** and remains almost constant despite other changes in the wave. The length of a wave (**L**) is equal to the product of the wave period (**P**) and the ***velocity*** of the wave (**V**):

$L = V \cdot P$

Long open-ocean waves or ***swells*** travel faster than short, locally generated sea waves. They also have longer wave periods and this is how they are distinguished from the short sea waves on reaching the coast. Long swells which have traveled hundreds of kilometres may have wave periods of up to 20 seconds. Smaller sea waves have wave periods of 5 to 8 seconds.

Where ocean depths are greater than the length of the waves, the wave motion does not extend to the ocean floor and therefore remains unaffected by the floor. As the ocean depth falls below half the wavelength, the wave motion becomes increasingly affected by the bottom. As the depth of water decreases the wave height increases rapidly and the wavelength decreases rapidly. Thus, the wave becomes more and more peaked as it approaches the shore, finally curling over as a ***breaker*** and breaking on the ***shore*** (**Figure 11m-1**). As the wave breaks, its potential energy is converted into kinetic energy, providing a large amount of energy for the wave to do work along the ***shoreline***. If you have ever watched waves breaking on a shore you may have observed that the waves appear to climb out of the water and also catch up to one another.

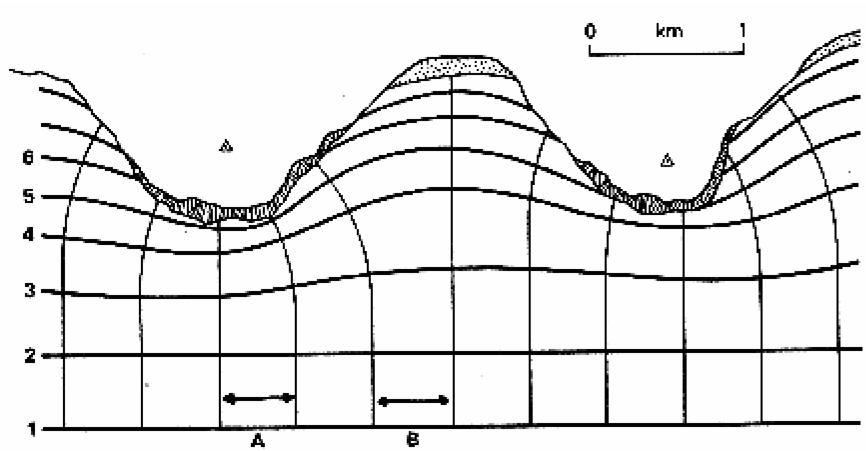
Figure 11m-1: Breaking waves on a beach.

Wave Refraction

Waves are subject to a reorientation, or ***wave refraction*** of their direction of travel as they approach the coast. Where oblique waves approach a straight shore, the frictional drag exerted by the sea floor turns the waves to break nearly parallel to the shore. On an indented coast the situation is more complex. The **animation** in ***Figure 11m-2*** shows an indented coast with a uniform underwater slope along its length. The wave crests are shown approaching the shore perpendicular to the general trend of the coast. The segments of the crests approaching the ***headlands*** begin to physically encounter the sea floor when they are just under a kilometer from the shore. They increase in height, decrease in wavelength, and slow down. The same crest approaching the ***bay*** continues unimpeded and so moves ahead of the wave segment off the headland. As a result of this process, headlands are usually sites of intense ***erosion*** while embayments are usually sites of sediment ***deposition***. Given enough time wave erosion will tend to create a smooth coastline.

Wave Refraction, Erosion, and Deposition

Segments A and B at position 1 in the **figure** below are in deep water and are unchanged. By the time they have reached position 3, A has slowed down and shortened its wavelength. It therefore lags behind B which is still unchanged. By the time the wave reaches position 5, A is about to break on the headland while B is advancing more slowly into the bay. The end result is that the crests try to conform to the outline of the shore and to break parallel to it. Segments A and B in deep water were the same width. The **orthogonals** which are drawn at right angles to the crests from the ends of segments A and B show that the length of A is shortened by about 30 percent at its breaking point (position 5), and B is lengthened to more than twice its deep water value at its breaking point. This means that the wave energy in segment A is concentrated onto the headland which causes wave height to increase in addition to the wave heightening caused by the shallowing of the water. Thus, since wave energy is proportional to wave height, the power of the waves is greater on the headland. In the bay, wave height is less since the energy of segment B is spread out. As a result, headlands are usually sites of intense ***erosion*** while embayments are usually sites of sediment ***deposition***. Given enough time wave erosion will tend to create a smooth coastline.



The following **photograph** shows the refraction of waves from above as they approach the shoreline.

Erosion, Transportation, and Deposition Along Coasts

A number of mechanical and chemical effects produce erosion of rocky shorelines by waves. Depending on the geology of the coastline, nature of wave attack, and long-term changes in sea-level as well as tidal ranges, erosional landforms such as ***wave-cut notches***, ***sea cliffs*** (**Figure 11m-3**) and even unusual landforms such as ***caves***, ***sea arches***, (**Figure 11m-4**) and ***sea stacks*** can form.

Figure 11m-3: Sea cliff along coast.

{PRIVATE**}Figure 11m-4:** Sea arches, Anse de l'Est, Ile aux Loups, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Transportation by waves and currents is necessary in order to move rock particles eroded from one part of a coastline to a place of deposition elsewhere. One of the most important transport mechanisms results from wave refraction. Since waves rarely break onto a shore at right angles, the upward movement of water onto the beach (***swash***) occurs at an oblique angle. However, the return of water (***backwash***) is at right angles to the beach, resulting in the net movement of beach material laterally. This movement is known as ***beach drift*** (see ***Figure 11m-5***). The endless cycle of swash and backwash and resulting beach drift can be observed on all beaches.

Frequently, backwash and ***rip currents*** cannot remove water from the shore zone as fast as it is piled up there by waves. As a result, there is a buildup of water that results in the lateral movement of water and sediment just offshore in a direction with the waves. The currents produced by the laterial movement of water are known as ***longshore currents***. The movement of sediment is known as ***longshore drift***, which is distinct from the beach drift described earlier which operates on land at the beach. The combined movement of sediment via longshore drift and beach drift is known as ***littoral drift***.

Tidal currents along coasts can also be effective in moving eroded material. While incoming and outgoing tides produce currents in opposite directions on a daily basis, the current in one direction is usually stronger than in the other resulting in a net one-way transport of sediment. Longshore drift, longshore currents, and tidal currents in combination determine the net direction of sediment transport and areas of deposition.

Many kinds of ***depositional landforms*** are possible along coasts depending on the configuration of the original coastline, direction of sediment transport, nature of the waves, and shape and steepness of the offshore underwater slope. Some common depositional forms are ***spits***, ***bayhead beaches***, ***barrier beaches*** or ***bay-mouth bars***, ***tombolos*** (**Figure 11m-6**), and ***cuspate forelands***.

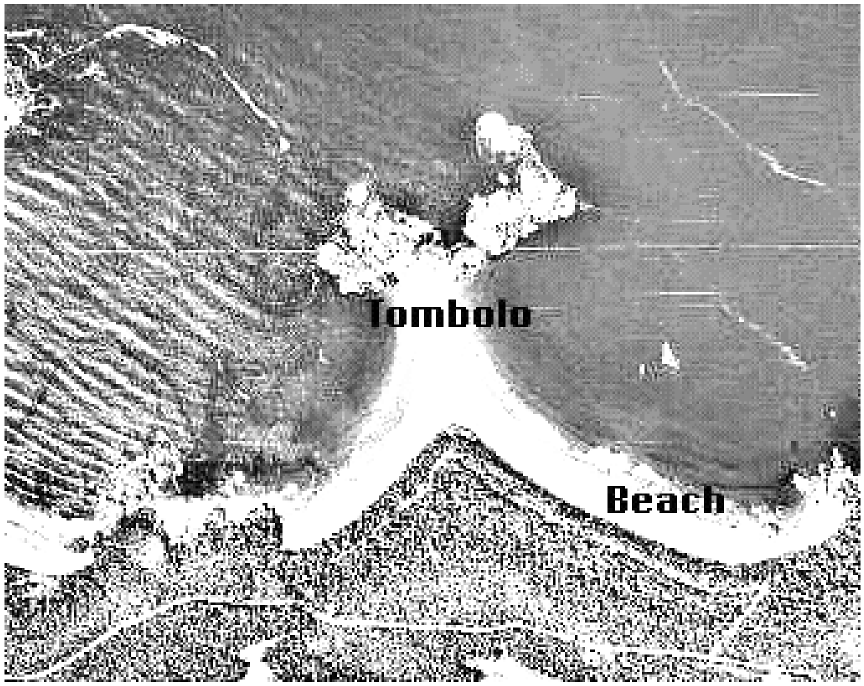


Figure 11m-6: Coastal features associated with erosion and deposition.

(n) Introduction to Glaciation

{PRIVATE}Various types of ***paleoclimatic*** evidence suggest that the climate of the Earth has varied over time. The data suggests that during most of the Earth's history, global temperatures were probably 8 to 15 degrees Celsius warmer than they are today. However, there were periods of times when the Earth's average global temperature became cold. Cold enough for the formation of ***alpine glaciers*** and ***continental glaciers*** that extended in to the higher, middle and sometimes lower latitudes. In the last billion years of Earth history, glacial periods have started at roughly 925, 800, 680, 450, 330, and 2 million years before present (B.P.). Of these ice ages, the most severe occurred at 800 million years ago when glaciers came within 5 degrees of the equator.

The last major glacial period began about 2,000,000 years B.P. and is commonly known as the ***Pleistocene*** or ***Ice Age***. During this glacial period, large glacial ice sheets covered much of North America, Europe, and Asia for long periods of time. The extent of the glacier ice during the Pleistocene, however, was not static. The Pleistocene had periods when the glaciers retreated (***interglacial***) because of mild temperatures, and advanced because of colder temperatures (***glacial***). Average global temperatures were probably 4 to 5 degrees Celsius colder than they are today at the peak of the Pleistocene. The most recent glacial retreat began about 14,000 years B.P. and is still going on. We call this period the ***Holocene epoch***.

In North America, the Pleistocene glaciers began their formation in the higher altitudes of the Rocky Mountains, and high latitude locations in Greenland and north-central Canada. From these locations, the ice spread in all directions following the topography of the landscape. In North America, the glaciers from the Rocky Mountains and north-central Canada met each other in the center of the continent creating an ice sheet that stretched from the Pacific to the Atlantic ocean. At their greatest extent, the ice sheets of North America covered most of Canada and extended into the United States to a latitude of about 40 degrees North.

A similar pattern of glaciation has also been scientifically documented in Europe and Asia. In Eurasia, ice sheets had their birth place in the Alps Mountains, Scandinavia, northern British Isles, and northern Siberia. The ice sheets of Eurasia, however, did not form a single ice sheet through convergence and their furthest extent south was limited to a latitude of about 45 degrees North.

Occurrence and Types of Glaciers

Today, glacial ice covers about 10 % of the Earth's land surface. During the height of the Pleistocene, ice sheets probably covered about 30 %. Currently, the most extensive continental glaciers are found in Antarctica and Greenland. We can also find smaller glaciers at higher elevations in various mountain ranges in the lower, middle, and higher latitudes.

Glaciers can be classified according to size. ***Continental glaciers*** are the largest, with surface coverage in the order of 5 million square kilometers. Antarctica is a good example of a continental glacier (**Figure 11n-1**).

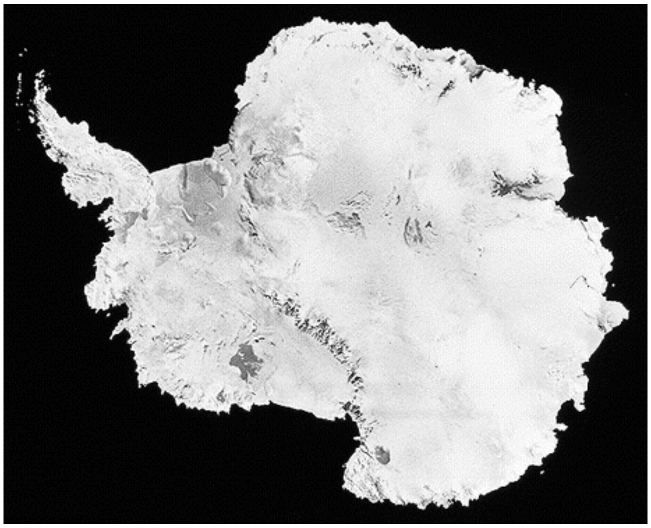


Figure 11n-1: Antarctica glacier (**Source:** *NASA*).

Mountain or ***alpine glaciers*** are the smallest type of glacier. These glaciers can range in size from a small mass of ice occupying a ***cirque*** to a much larger system filling a mountain valley (**Figure 11n-2**). Some mountain glaciers are even found in the tropics. The merger of many alpine glaciers creates the third type of glacier, ***piedmont glaciers*** (**Figure 11n-3**). Piedmont glaciers are between several thousand to several tens of thousands of square kilometers in size.

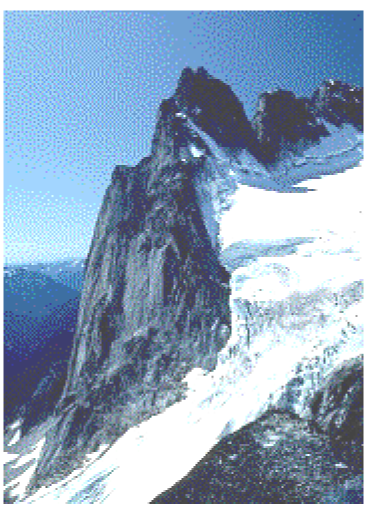


Figure 11n-2: Small alpine valley glacier.

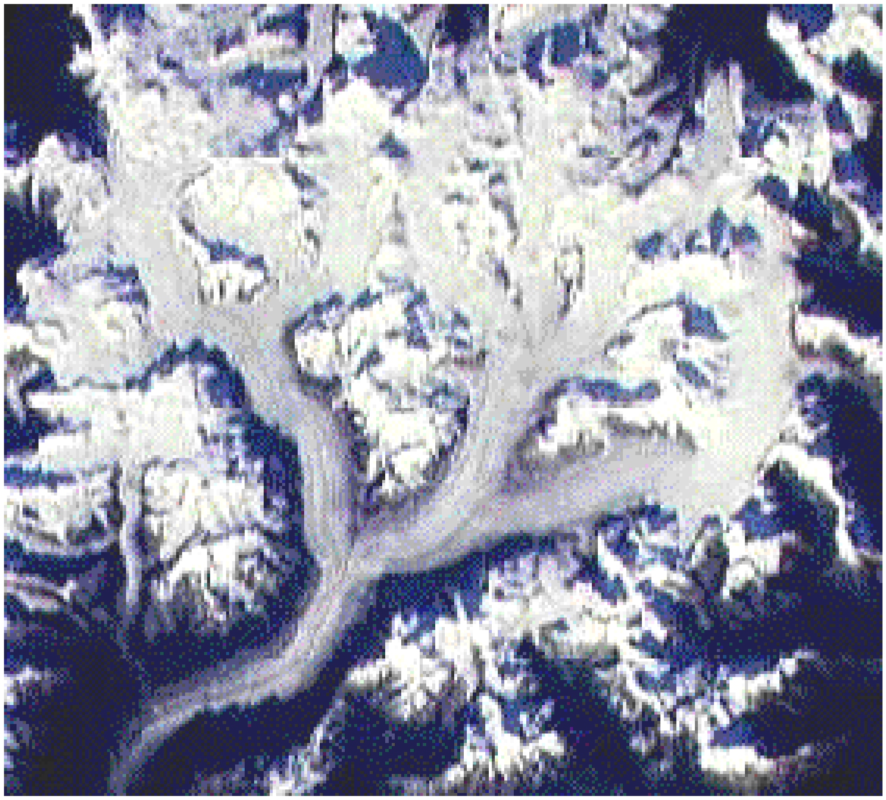


Figure 11n-3: Merging alpine glaciers as seen from above (**Source:** *NASA*).

(o) Glacial Processes

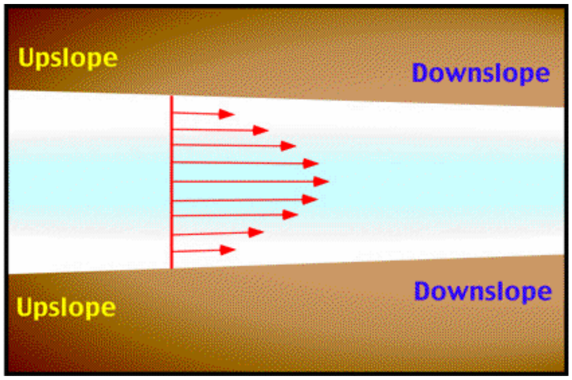
{PRIVATE}Growth of Glaciers

Ice that makes up ***glaciers*** originally fell on its surface as ***snow***. To become ice, this snow underwent modifications that caused it to become more compact and dense. Glacial ice has a density of about 850 kilograms per cubic meter. The density of snow ranges from about 50 to 300 kilograms per cubic meter (the density of fresh water is approximately 1000 kilograms per cubic meter). After the snow falls, the crystals can be reduced by the effects of ***melting*** and ***sublimation***. Scientists call this process ***ablation***. For most glaciers, ablation is a phenomena dominant in the summer months. The snow also undergoes physical compaction through melting and refreezing. At first, these processes cause the original snowflakes to be transformed into small round crystals. This partly melted, compressed snow is called ***névé***. ***Névé*** has a density exceeding 500 kilograms per cubic meter. If the névé survives the ablation that occurs during the summer months it is called ***firn***. When this process happens year after year, a number of layers of firn can accumulate. Accumulation then causes a further increase in density, modifying the firn into glacier ice, as the lower layers of firn are compressed by the weight of the layers above. On average, the transformation of névé into glacial ice may take 25 to 100 years.

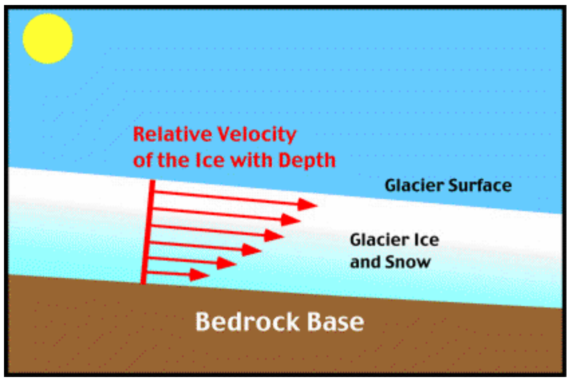
Glacier Movement

To be called a glacier, a mass of ice must be capable of motion. Glacial movement occurs when the growing ice mass becomes too heavy to maintain its rigid shape and begins to flow by plastic deformation. In most mountain glaciers, flow of ice begins with accumulations of snow and ice greater than 20 meters.

Flow rates within the various regions of a glacier are not uniform. From directly above, the middle of the glacier appears to flow with the greatest speed (**Figure 11o-1**). At the margins of glacier, surface movement is slowed down because of the frictional effects of the valley wall. Looking at the glacier in cross-section, we notice that the bottom of the glacier also moves slowly, once again, because of the influence of frictional forces (**Figure 11o-2**).

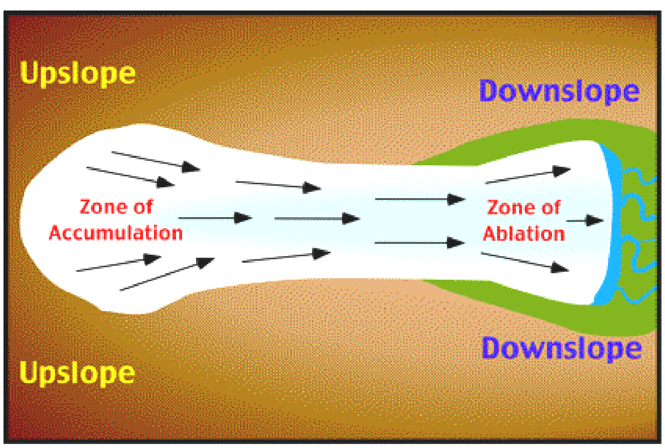


{PRIVATE**}Figure 11o-1:** Overhead view of an alpine valley glacier showing the relative speed of ice movement.



{PRIVATE**}Figure 11o-2:** Long cross-section view of an alpine valley glacier showing the relative speed of ice movement.

Figure 11o-3 illustrates the typical pattern of ice flow in a valley glacier. In the upper reaches of the glacier, ice and snow accumulate in a broad basin formed by the effects of ***physical weathering*** and ***erosion***. As the glacier proceeds to move downslope, the flow lines of the ice begin to converge because of the narrowing of the valley. This convergence causes a compression of the ice flow in central section of the glacier. At the terminal end of the glacier, flow lines spread out as the ice is no long constricted by valley walls.



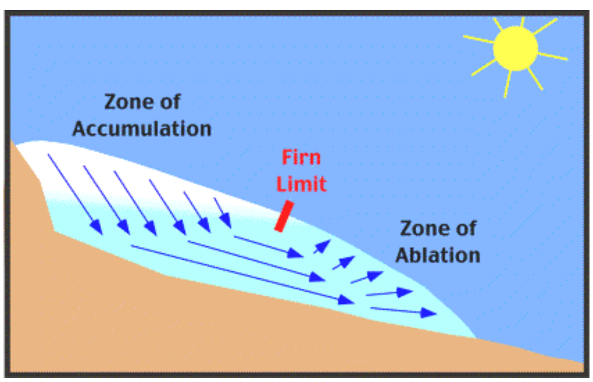
{PRIVATE**}Figure 11o-3:** Diagram showing ice flow patterns along the entire length of an idealized alpine glacier.

The velocity of flow of glacier ice is influenced by a variety of factors. Some of the more important factors are the gradient of the valley floor, the temperature and thickness of the ice, and the constriction caused by the valley walls. The movement of ice over the ground in most ***temperate glaciers*** is enhanced by a process known as ***basal sliding***. The immense pressure caused by the weight of the overlying glacial mass causes the ice making contact with the ground to melt because of pressure, despite subzero temperatures (***pressure melting***). The melting ice then forms a layer of water that reduces the friction between the glacial ice and the ground surface. This water then facilitates the movement of the ice over the ground surface by producing a layer with very little friction. Because of basal sliding, some glaciers can move up to 50 meters in one day. However, average rates of movement are usually less than 1 meter per day. ***Cold glaciers*** tend to move very slowly because there is no basal sliding. Movement in these glaciers takes place mainly due to internal slippage of the ice over the ground.

Glacier Mass Balance

Scientists often view glaciers as ***systems*** that are influenced by a number of ***inputs*** and ***outputs***. The main inputs to the glacial system are water, in the form of ***snow***, and ***eroded*** sediments that are picked up by the moving ice. Water leaves the glacial system when ice is converted into water or vapor. Sediment is deposited at the base of the glacier as ***till*** and at its terminal end as ***moraines*** or materials reworked by ***glaciofluvial*** processes.

We can also use a systems approach to help us understand why glaciers expand and shrink, and advance and ***retreat***. This type of modeling is referred to as glacier ***mass balance***. The mass balance of a glacier involves two main components: ***accumulation*** of snow in the glacier's ***zone of accumulation*** and the ***ablation*** of ice in the ***zone of ablation*** (**Figure 11o-4**). The zone of accumulation occurs in the upper reaches of the glacier where yearly additions of snow exceed losses due to ***melting***, ***evaporation***, and ***sublimation***. The surface of this zone is covered by snow throughout the year. Below the zone of accumulation is the zone of ablation. In this zone, the losses of snow and ice from melting, evaporation, and sublimation are greater than the additions. The line that separates these two areas is called the ***firn limit*** or ***snow line***.



{PRIVATE**}Figure 11o-4:** The following illustration describes the major components of glacier mass balance. Additions to the glacial system occur in the ***zone of accumulation*** where snow is converted into glacial ice over time. This ice then flows downslope into the zone of ablation. In the ***zone of ablation***, losses occur from the glacier from the ***melting***, ***evaporation***, and ***sublimation*** of solid and liquid forms of water. The ***firn limit*** marks the separation point between the two zones. Above the firn limit snow is able to survive the summer season.

Forward flow of glacial ice is controlled by gravity and the ***accumulation*** of snow in the ***zone of accumulation***. If losses due to ***ablation*** are identical to accumulations, the glacier will appear to be standing still in spite of the fact that the ice is actually moving forward. Advance of the glacier's ***terminus*** occurs when net accumulation is greater than net ablation. **{**PRIVATE**}**Advance of the glacier's ***terminus*** occurs when the addition of snow in the zone of accumulation exceeds losses of ice in the ***zone of ablation***, the ***firn line*** moves forward as the ***zone of accumulation*** expands with increasing snowfall. Glacial retreat takes place when net accumulation is less than net ablation, {PRIVATE}when the addition of snow in the ***zone of accumulation*** is less than the losses of ice in the ***zone of ablation,*** the ***firn line*** moves upslope as the zone of accumulation shrinks with decreasing snowfall.

Ideally, the rate of flow and movement of the glacier's terminus should be controlled by net ccumulation and net ablation. This relationship, however, is not temporally immediated. In many glaciers, there are significant time lags between one year's net accumulation and net ablation and the corresponding movement of the glacier. Sometimes this lag can be in the order of several decades when the glaciers are quite large. Some glaciers can experience a rapid forward ***surge*** as great as 10 to 20 meters per day. This phenomenon is believed to be caused by large inputs of snow sometime in the past.

Eventually, all glacier ice is lost in the zone of ablation by the processes of ***melting***, ***evaporation***, and ***sublimation***. Another process that can remove mass from a glacier is ***calving***. This process occurs in glaciers whose terminus reaches large bodies of water. Calving involves the separation of portions of the glacier ice into the water body. Many icebergs enter the oceans of the world from the calving of the Greenland and Antarctic glaciers.

Today most glaciers are retreating because of the general warming of global temperatures since the beginning of this century. This indicates that the mass balances of these glaciers are negative because of less snow accumulating or higher levels of ablation. During the ***Little Ice Age***, when global temperatures were cooler than present, many glaciers over much of the world made strong advances.

(p) Landforms of Glaciation

Glaciers have played an important role in the shaping of landscapes in the middle and high latitudes and in alpine environments. Their ability to ***erode soil*** and ***rock***, ***transport sediment***, and ***deposit sediment*** is extraordinary. During the last ***glacial period*** more than 50 million square kilometers of land surface were geomorphically influenced by the presence of glaciers.

Glacial Erosion

Two major ***erosional*** processes occur at the base of a ***glacier***. First, at the base of a glacier, large amounts of loose ***rock*** and ***sediment*** are incorporated into the moving glacial ice by partial melting and refreezing. The second process of erosion involves the ***abrasive*** action of the held rock and sediment held by the ice on the surface underneath the glacier. This abrasive process is known as ***scouring***. Scouring creates a variety of features. The most conspicuous feature of scouring is ***striations*** (**Figure 11p-1**). Striations appear as scratches of various size on rock surfaces. In some cases, abrasion can polish the surface of some rock types smooth. This geomorphic feature is known as ***glacial polish***. The abrasive action of scouring also produces a fine ***clay***-sized ***sediment*** that is often transported away from the glacier by ***meltwater***. As a result of this process, glacial meltwater can have a light, cloudy appearance, and is called ***glacial milk***.

{PRIVATE}**Figure 11p-1:** Glacial striations, Lac Blanchet, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

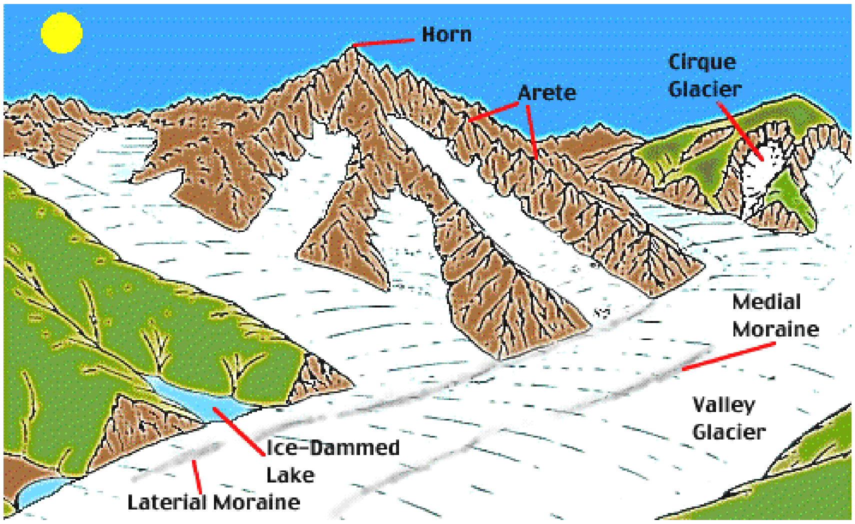
The second major ***erosional*** process that occurs at the base of a ***glacier*** is ***plucking***. Plucking is the process of particle detachment by moving glacial ice. In this process, basal ice freezes in rock surface cracks. As the main body of the glacial ice moves material around the ice in the cracks is pulled and plucked out. The intensity of the plucking process is greatest on the ***lee-side*** of rock mounds. When combined with glacial abrasion, the action of plucking on rock mounds produces a unique asymmetrical feature known as ***roche moutonnee***. Roche moutonnee are smooth on the side of ice advancement and steep and jagged on the opposite side.

Glaciers generally flow over the land surface along a path of least resistance. The flow of an ***alpine glacier*** into a valley, causes the glacier to rapidly advance producing a swollen tongue of ice at the glacier's ***snout***, known as a ***lobe***. As the lobe moves down the valley it often encounters the lobes of other glaciers from connecting valleys. The glacier grows in size with addition of the flow of connected sub-valleys. The following **image** illustrates one of these networks of connected alpine glaciers (**Figure 11p-2**).

Figure 11p-2: Merging alpine glaciers viewed from above (**Source:** *NASA*).

A number of distinct ***erosional*** features can be observed in mountainous regions that have experienced the effects of glaciation. Much of this erosion is exerted on the bottoms and sides of alpine valleys that guide the flow of glaciers. This erosion causes the bottom and the sides of any glaciated valley to become both wider and deeper over time. Glacial erosion also results in a change in the valley's cross-sectional shape. Glacial valleys tend to have a pronounced U-shape that contrasts sharply with V-shape valleys created by stream erosion. Small adjoining feeder valleys entering a large valley in a glaciated mountainous region tend to have their floors elevated some distance above the level of the main valley's floor. Geomorphologist call this landform a ***hanging valley***. Hanging valleys develop because larger, more massive glaciers create more erosion and deeper valleys. Many hanging valleys are also the sites of sensational waterfalls.

Some of the other features associated with glacier erosion in alpine regions are ***cirques***, ***horns***, and ***arêtes*** (**Figure 11p-3**). ***Cirques*** are the bowl shaped depressions found at the head of glacial valleys. For most alpine glaciers, cirques are the areas in the alpine valleys where snow first accumulated and was modified into glacial ice. The glaciers that occupy cirques are called ***cirque glaciers***. ***Horns*** are pyramidal peaks that form when several cirques chisel a mountain from three or more sides. The most famous horn is the Matterhorn found in the Swiss Alps. ***Arêtes*** are the narrow serrated ridges found in glaciated alpine areas. Arêtes form when two opposing cirques back erode a mountain ridge.



{PRIVATE}**Figure 11p-3:** Features associated with alpine glaciation.

Talus and other foot-slope deposits are also common in a glaciated valley. Because of the enhancement of ***freeze-thaw*** processes bedrock in alpine areas is ***weathered*** by the growth of ice crystals. This type of weathering shatters the bedrock into sharp angular fragments that accumulate at the bottom of rock slopes as talus. Much of the debris carried by an alpine glacier comes from valley sides where talus accumulates.

The ***erosional landforms*** produced by ***continental glaciers*** are usually less obvious than those created by alpine glaciers. Like alpine glaciers, the movement of continental glaciers followed topographic trends found in the landscape. Continental ice sheets were very thick, between 1000 to 3000 meters. The mass of these glaciers covered all but the highest features and had extremely strong erosive power. Much of the ***Canadian Shield*** shows the effects of abrasion and gouging which created ***glacial polish*** and ***striations*** on bedrock surfaces. In some areas, continental ice sheets produced huge U-shaped valleys from previously V-shaped stream valleys. In other areas, erosion by the continental ice sheets scooped out large shallow basins, many of which exist today as lakes. Many of the lakes on the Canadian Shield, including those of the Great Lakes, were created by glacial erosion.

Glacial Deposition

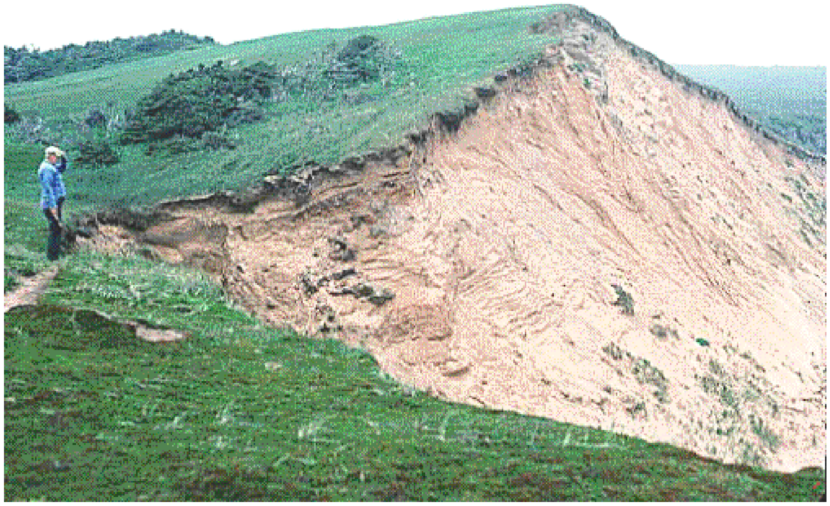
A large part of the surface of a glacier is covered with a coating of sediment and rock debris. This is especially prevalent near the snout of the glacier, where most of the ice has been lost to ablation and sediment is left behind. Sediment is added to glacial ice in two ways. Large quantities of sediment are picked up by ***abrasion*** and ***plucking*** at the base of the ice. In alpine areas, sediment is added to the surface of the glacier from the valley walls through various types of ***mass movement***. Much of the debris that is added to the ice of the glacier is eventually delivered to the snout because of the continual forward flow of glacial ice. From the snout this material can be placed directly from the ice or it can be deposited through the action of flowing ***meltwater***. Geomorphologists call the later deposits ***glaciofluvial*** deposits. The technical term used to describe material deposited by the ice is called ***till*** or ***moraine***. All glacial deposits are by and large known as ***glacial drift***.

Till is a heterogeneous combination of unstratified sediments ranging in size from large ***boulders*** to minute particles of ***clay***. When till is deposited along the edge of a glacier it tends to form irregular hills and mounds known as ***moraines***. A ***terminal moraine*** is a deposit that mark, the farthest advance of a glacier. Moraine deposits created during halts in the ***retreat*** of the glacier are called ***recessional moraines***. The debris that falls from valley side slopes can be concentrated in a narrow belt and cause a deposit known as a ***lateral moraine*** (**Figure 11p-3**). When two glaciers flow together, two lateral moraines can merge to form an interior belt of debris, called a ***medial moraine*** (**Figure 11p-3**). A ***till plain*** is a large, relatively flat plain of till that forms when a sheet of ice becomes detached from the main body of the glacier and melts in place. Sometimes the sediments in a till plain can contain large boulders. If these ***boulders*** are transported a great distance from their place of origin, they are called ***erratics*** (**Figure 11p-4**).

{PRIVATE**}Figure 11p-4:** Glacial erratic near Point Lake, Northwest Territories. Glacial erratics are large pieces of rock that have been transported away from their source areas by moving glacial ice sheets. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Glaciofluvial deposits are generally quite stratified and less assorted in particle size. ***Outwash*** deposits are formed when sand is eroded, transported, and deposited by meltwater streams from the glacier's snout and nearby till deposits to areas in front of the glacier. ***Outwash plain*** develops when there are a great number of meltwater streams depositing material ahead of the glacier (**Figure 11p-5**). Glaciofluvial deposits are also directly in front of the glacier. Where water rich in sediment flows off the snout of the ice, a conical-shaped pile of sediment, known as a ***kame***, can be deposited (**Figure 11p-6**). Many kames are often found on or at the edge of moraines.

{PRIVATE**}Figure 11p-5:** Glacier snout and outwash plain, Bylot Island, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).



{PRIVATE**}Figure 11p-6:** Kame, La Bluff, Ile de la Grande Entree, Canada. This kame was deposited during the main stage of the last glaciation when the Laurentide ice sheet filled the Gulf of St. lawrence. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Glaciers can also contain sinuous flows of meltwater that occur in ice tunnels at the base of the ice. The beds of these sub-surface glacial streams are composed of layers of ***sand*** and ***gravel***. When the ice melts from around the meltwater tunnels, the beds of sand and gravel are deposited on the Earth's surface as long twisting ridges known as ***eskers*** (**Figure 11p-7**).

{PRIVATE**}Figure 11p-7:** Esker near Lac du Sauvage, Northwest Territories. The slightly curving thin ridge in the centers of this photo is the esker. The flat region in the foreground to the left of the esker was formed by glacial outwash. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

When glaciers are rapidly ***retreating***, numerous blocks of ice can become detached from the main body of the glacier. If glacial drift is then placed around the ice, a depression on the surface called a ***kettle hole*** can be created when the ice melts . Kettle holes are commonly found on moraine and outwash plain deposits. Large kettle holes that reach below the water table can form into lakes. The **photograph** below shows some kettle lakes in glaciofluvial outwash complex located in the Northwestern District of Mackenzie, Northwest Territories (**Figure 11p-8**). Some kettle holes develop into wetlands such as bogs, swamps, and marshes.

{PRIVATE**}Figure 11p-8:** Kettle Lakes, Northwest Territory, Canada. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Glacial retreat also creates hill shaped deposits of till known as a ***drumlins*** (**Figure 11p-9**). Drumlins often occur in large congregations across areas of New York and Wisconsin, USA and Ontario, Canada. The streamline shape of these glacial features resembles a protracted teaspoon laying bowl down. The narrow end of the drumlin points to the general direction of glacial retreat. Drumlins also come in a variety of dimensions. Lengths can range from 100 to 5000 meters and heights can sometimes exceed 200 meters.

{PRIVATE**}Figure 11p-9:** Drumlin field in northwestern Manitoba. These features are made of till and are formed at the base of a retreating glacial ice sheet. The long axis of feature gives the orietation of glacial movement. The narrow end points in the direction of retreat. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

(q) Periglacial Processes and Landforms

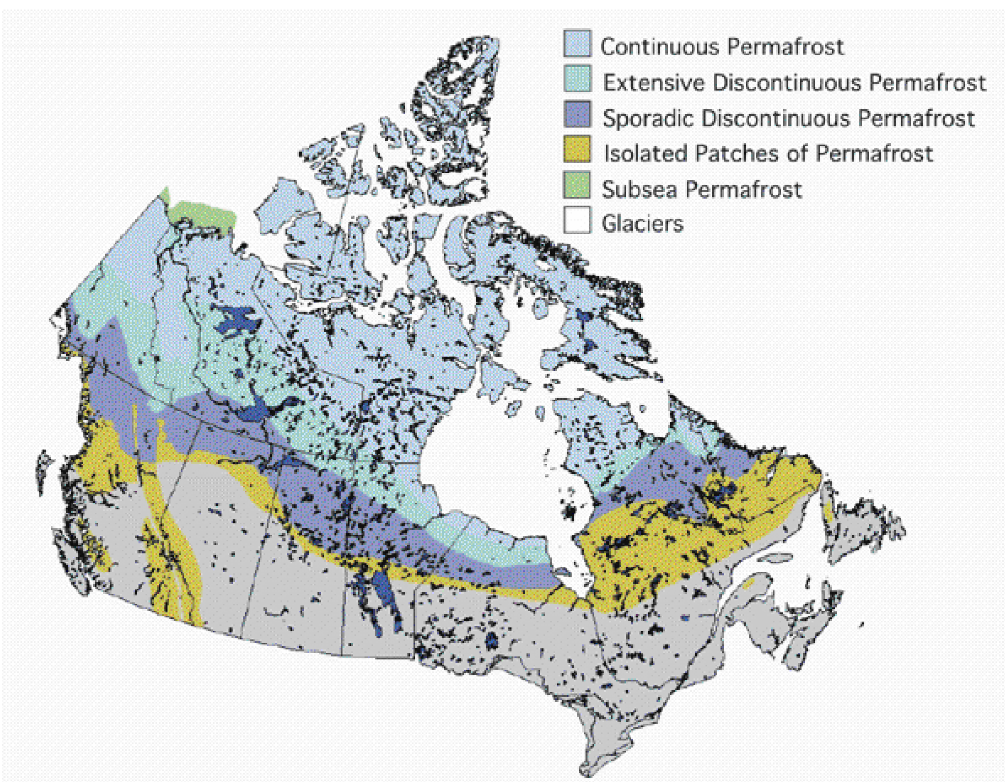
{PRIVATE}Introduction

Several definitions exist for the term ***periglacial***. The earliest definitions suggested that these geomorphic environments were located at periphery of past ***Pleistocene*** glaciers. In these environments, the landscape is dominantly influenced by frost action. However, frost action also influences landscapes that were not at the margin of ancient glaciers. For this reason, we should use a broader definition of this term. This definition suggests that in a periglacial environment the effects of ***freezing*** and ***thawing*** drastically modify the ground surface. Types of modification include the displacement of soil materials, migration of groundwater, and the formation of unique landforms. More than a third of the Earth's terrestrial surface can be included in this definition.

Permafrost

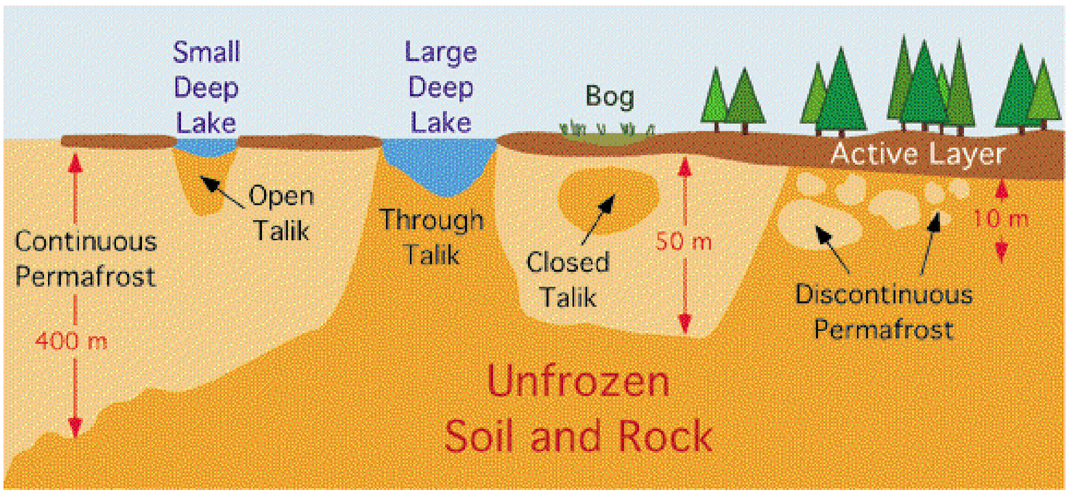
Permafrost is a condition where a layer of soil, sediment, or rock below the ground surface remains frozen for a period greater than a year. Permafrost is not a necessary condition for creating periglacial landforms. However, many periglacial regions are underlain by permafrost and it influences geomorphic processes acting in this region of the world.

Permafrost is found in about 25 percent of the Earth's non-glaciated land surface (**Figures 11q-1** and **11q-2**). It occurs wherever the mean annual ground temperature is less than 0 degrees Celsius. The southern limit in Canada generally correlates with an annual surface air temperature of -8 degrees Celsius. In some areas, permafrost can be up to 1500 meters deep. Common depths of permafrost are several hundred meters. Most deposits of permafrost have an upper ***active layer*** that is between 1 to 3 meters thick. This active layer is subject to a cyclic thaw during the summer season.



{PRIVATE}Figure 11q-1: Distribution of the various types of permafrost in Canada. More than 50 % of Canada is covered by some form of permafrost. This map does not identify areas of alpine permafrost (**Source:** *Natural Resources Canada - Terrain Sciences Division - National Permafrost Database*).

In some places, localized unfrozen layers or ***taliks*** are located on top, underneath, or within masses of permafrost (**Figure 11q-2**). Often in continuous permafrost areas, taliks are found under lakes because of the ability of water to store and vertically transfer heat energy. The vertical extent of the taliks found under lakes is related to the depth and volume of the overlying water body. Larger water bodies can store and transfer more heat energy downward. A ***closed talik*** is unfrozen ground that is found within a mass of permafrost. Closed taliks can develop when lakes fill in with sediment and become bogs. With the removal of the open water, summer solar radiation is now being received by a surface with a lower ***specific heat*** and poor vertical heat transfer. As a result, soil near the surface begins to freeze solid encasing a zone of unfrozen soil in permafrost. Closed taliks can also form because of groundwater flow.



{PRIVATE}**Figure 11q-2:** Vertical cross section of the transition zone between continuous and discontinuous permafrost. The graphic also shows the various types of talik or unfrozen ground. ***Open talik*** is an area of unfrozen ground that is open to the ground surface but otherwise enclosed in permafrost. ***Through talik*** is unfrozen ground that is exposed to the ground surface and to a larger mass of unfrozen ground beneath it. Unfrozen ground encased in permafrost is known as a ***closed talik***.

Five different types of permafrost have been recognized (**Figures 11q-1** and **11q-2**). ***Continuous permafrost*** exists over the landscape as an uninterrupted layer. Zones of permafrost with numerous scattered small thawed areas are called ***discontinuous permafrost***. Discontinuous permafrost in usually exists as an extensive marginal zone at the edge of continuous permafrost. ***Sporadic permafrost*** occurs where small islands of permafrost are scattered in generally unfrozen areas. ***Alpine permafrost*** is found at higher elevations in areas outside where continuous or discontinuous permafrost is common. ***Subsea permafrost*** is frozen ground that exist in the sediments beneath seawater. Large areas of subsea permafrost exist along the northern coastal edge of Russia, Alaska, and parts of northern Canada.

Periglacial Processes: Weathering

Locations that have a periglacial environment are characterized by the presence of large quantities of angular, fractured rock (**Figure 11q-3**). The angular nature of these deposits suggests that the process responsible for the rock fracturing is the crystallization of water. The quantity of the deposits indicates that the frost weathering process operates over and over again in repeated cycles of ***freeze-thaw***. In fact, repeated thawing allows further fracturing because the liquid water is able to fill newly developed cracks.

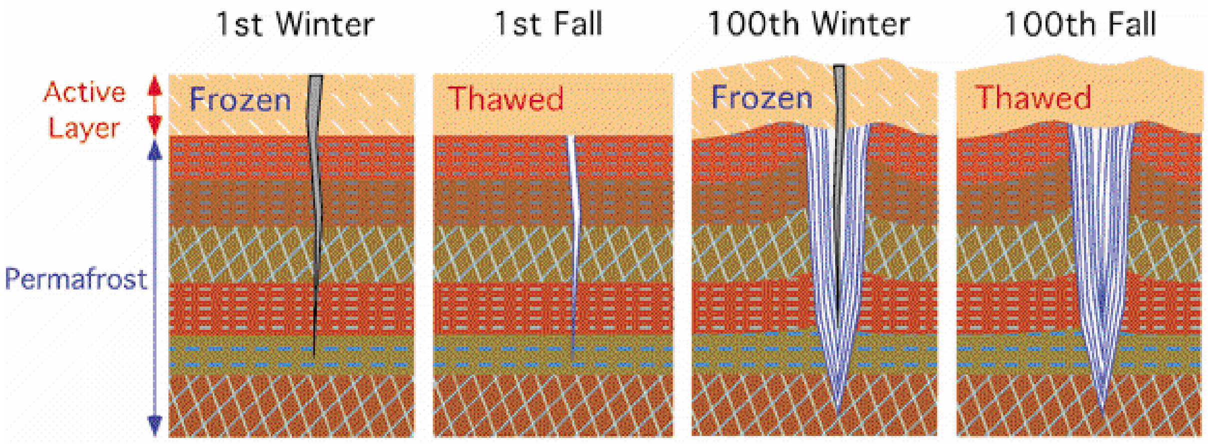
{PRIVATE**}Figure 11q-3:** Frost-shattered granite bedrock (**felsenmeer**), northern Manitoba. This is a close-up of frost-shattered bedrock, consisting of angular blocks of Precambrian granite. Extensive areas of blocks are called felsenmeer. The shattering has occurred in permafrost terrain, near treeline. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

The critical temperature for the development of frost induced fracturing is between -4 and -15 degrees Celsius. Scientists have also discovered that relative slow rates of cooling produce the greatest quantity of fracturing stress. Measurements for marble and granite indicate that a rate of cooling between -0.1 and -0.5 degrees per hour is optimal.

Periglacial Processes: Ground Ice

Surface soils and sediments in periglacial environment are frequently influenced by a variety of different types of ***ground ice***. Some of these masses of ground ice can be as much as 30 meters across. The most common form of ground ice is ***pore ice***. Pore ice develops in the pore spaces between soil and sediment particles where liquid water can accumulate and freeze. Another common form of ground ice is ***needle ice***. Needle ice consists of groups of narrow ice slivers that are up to several centimeters long. They normally form in moist soils when temperatures drop below freezing overnight. Needle ice plays an active role in loosening soil for ***erosion*** and tends to move small rocks upward to the soil surface. On sloped surfaces, needle ice can also enhance ***soil creep*** by moving soil particles at right angles to the grade.

Ice wedges are downward narrowing masses of ice that are between 2 to 3 meters wide at the base and extend below the ground surface up to 10 meters. It is believed that they form when a seasonal crack in the ground forms in the winter. During this season, extreme cold temperatures can cause soil contraction. Soil and ice, like other solid materials, contracts as temperatures drop way below 0 degrees Celsius. At first, the crack is several millimeters wide and about a meter deep. When temperatures warm up in the summer, liquid water from the active layer fills the crack. This water then refreezes because the fracture extends into the sub-zero permafrost. The freezing of the water results in a volumetric expansion of about 9 %. This expansion then increases the width and depth of the fracture. At the same time, the soil layers adjacent to the developing ice wedge are pushed outward and upward creating a slight hill at the surface. The processes just described cause the ice wedge to increase in size with each passing year. Each winter new cracks form in the ice wedge because of contraction. In the summer, more liquid water seeps in to the cracks to become frozen and the summer warming results in the surface mound getting bigger (**Figure 11q-4**). Each cycle of ice addition can be identified as narrow layers called foliations. In arid cold environments, some ice wedges can accumulate quantities of wind blown sand within their winter cracks. These particular types of ground ice are often given the special name ***sand wedges***.



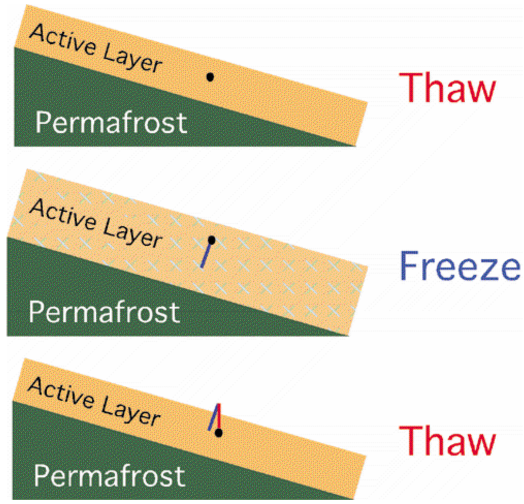
{PRIVATE**}Figure 11q-4:** The following series of graphics showing the evolution of ice wedges. The development of an ice wedge begins with cold temperatures in winter forming a crack in the ground surface that is about a meter deep and a few millimeters wide. In the warmer months, melting of the active layer liberates liquid water which flows into the crack. This water then freezes on contact with the permafrost. The cycle just described repeats itself year after year with new cracks forming in the developing ice wedge. The two graphics on the right show a well developed ice wedge after many cycles of development.

Segregated ice are masses of almost pure ice that grow within permafrost because of liquid water diffusion. Many forms of segregated ice exist as extensive horizontal layers. In this form, they are sometimes called ***ice lenses***. In the development of these features, water migrates under some type of pressure from unfrozen parts of the soil to the ice mass. When the liquid water comes in contact with the segregated ice mass it freezes enlarging the feature. Water diffusion can be caused by temperature and pressure gradients, ***gravity***, and by the movement of groundwater under pressure. Most masses of segregated ice are found just below the active layer. The pressure exerted by overlying sediments controls the growth of segregated ice.

Periglacial Processes: Mass Movement

During the warmer seasons, ***mass movement*** can be a common phenomenon in periglacial environments. It usually occurs in four forms: ***solifluction***; ***gelifluction***; ***frost creep*** and ***rockfalls***. ***Solifluction*** is the slow downslope flow of soil and sediment that is saturated with water. This process can occur on very shallow grades. The common sign of this form of mass movement is the presence solifluction lobes, tongue like semi-mixed surface deposits. In periglacial environments, solifluction is confined to times when temperatures are well above zero and free liquid water is available in the active layer. Solifluction is very common when surface sediments are poorly drained and quite saturated with water. ***Gelifluction*** is a form of solifluction where the moving materials slide over a slick permafrost layer.

Frost creep is the slow downslope movement of soil and sediment because of frost heaving and thawing (**Figure 11q-5**). The process begins with the freezing of the ground surface elevating particles at right angles to the slope. The particles are elevated because cold temperatures causes water in between particles to freeze and expand. In the warm season, thawing causes the ice to convert back to liquid water and the contracting surface drops the particles in elevation. This drop, however, is influenced by ***gravity*** causing the particle to move slightly downslope.



{PRIVATE**}Figure 11q-5:** The following series of graphics show how frost creep moves soil particles gradually downslope. The top graphic shows the position of a single particle in a thawed active layer above a zone of permafrost. When cold temperatures occur the active layer freezes. This freezing causes the water between the soil particles to expand making the layer thicker. The expansion processes also moves the illustrated soil particle up. This movement occurs perpendicular to the slope. With the return of warmer temperatures, the active layer contracts as the ice between soil particles melts. However, this contraction is controlled by the force of gravity causing the illustrated soil particle to move forward as it drops.

Exposed rock outcrops are common in periglacial regions of the world. Cold temperatures experienced here do not favor the development of an extensive layer of soil and regolith over bedrock. As a result, exposed rock on slopes is subject to this region's harsh environment. Extreme variations in temperature can cause the fracturing of rock along natural bedding planes and joints leading to ***rockfalls***. This fracturing is mainly due to ***frost*** and ***insolation weathering*** (see ***Lecture 4b***).

Periglacial Processes: Erosion

Processes of erosion and deposition in periglacial parts of the world tend to have their own unique character. These characteristics are related to the importance of freeze-thaw action, the presence of strong winds, and the fact that the warm season is very short. As a result, three forms of processes tend to dominate in periglacial environments: ***nivation***; ***eolian erosion*** and ***deposition***; and ***fluvial erosion*** and ***deposition***.

Nivation is the localized form of erosion associated with isolated patches of snow that remain through the summer season. It involves three separate processes: frost weathering at the margins of snow patches; meltwater erosion; and gelifluction. Snow patches that persist through the warm season create an environment for physical weathering at their margins. Temperatures at the margins of the snow fluctuate between above and below zero degrees Celsius diurnally. As a result, water in the cracks of rocks located by the snow changes from liquid to solid many times, quickly creating a mass of small fragments. If the snow patch is on a grade, meltwater along the base of the snow patch will transport the weathered rock fragments downslope. The snow patches can also insolate the soil and sediments that lay underneath it. The effects of insolation can raise the temperature of these materials to above freezing when combined with seeping water from melted snow. Once thawed and saturated, the sediments may begin moving by ***gelifluction*** if a slope exists.

A common feature created by nivation is a ***nivation hollow***. These features have been known to develop under snow patches in just a few seasons. The usual location of these features in the Northern Hemisphere is on a slope that faces northeast. This orientation is of course most protected from the warming effects of sun's rays. The development of the hollows requires two ingredients: a snow patch that returns to the same area year after year and a slope to allow for erosional transport of material out of the developing depression. The process begins with a patch of snow. Around the edges of this patch, physical weathering and frost heaving begins to separate particles for erosion. Running water then picks up the loose particles and carries them off. Material is also removed from the developing hollow by gelifluction. As the summer season progresses, the patch of snow reduces in size and the excavation of material continues inward. Enlargement of the hollow involves several different mechanisms. Sometime in the following year, the boundary of the snow mound and depression will once again be inline and frost weathering will eat away at the hollow's edge. The edge of the hollow is also preferentially eroded because the micro-slope creates localized instabilities and focuses the entrainment potential of flowing water.

Strong winds are a characteristic of periglacial areas. These winds often move large quantities of loose sediment and soil. This is especially true during the summer months when increases in stream discharge, drying of sediments, and the melting of snow and ice make more material available for eolian transport. Some of the depositional features that are created include ***sand dunes***, ***loess*** accumulations, and ***sand sheets***.

The ***discharge*** of most streams in periglacial regions exhibits particular temporal patterns. Most of the flow tends to occur during a period of weeks when snowmelt occurs. The discharge patterns of these streams may also show wide fluctuations that are timed diurnally. This is unlike temperate and tropical streams where discharge is taking place throughout the year. The concentration of discharge in a short period leaves some peculiar features. These short-lived streams tend to have poorly developed shallow braided channels. When discharge is suddenly reduced, large quantities of gravel and boulders are left on the landscape where the flow was taking place. The final unique characteristic of periglacial streams is that their channels can be beaded with deep pools. Beaded channels develop when a stream passes over a network of ice wedges. Thermal properties of the flowing water cause the ice wedges to melt producing pools.

Periglacial Landforms

The surface of periglacial areas is often characterized by the presence of ground materials arranged in a variety of symmetrical, geometric shapes (**Figures 11q-6** and **11q-7**). These features are collectively known as ***patterned ground***. Shapes of patterned ground can include stripes, steps, circles, polygons, and nets. Sometimes one pattern can morph into another shape. Researchers have discovered that a single process cannot explain the various forms observed. Many of these features appear to be caused by freeze-thaw action selectively moving coarse particles to the edge of the shape or to its surface. Some polygon forms seem to be caused by the same thermal processes that create ice wedges. Yet, other types of pattern ground still remain not fully explained.

{PRIVATE**}Figure 11q-6:** Stone circles on Melville Island, Northwest Territories, Canada. The patterns on the ground consist of circular arrangements of slabs (foreground) and small domes of mud with stony rims. The features were produced by the churning action of frost forming in unsorted glacial till and regolith. Countless freeze-thaw cycles sorted the surface debris, continually heaving the finer matter to the surface, and leaving the coarser fragments around the edges. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

{PRIVATE**}Figure 11q-7:** Ice-wedge polygons in peatland, Hudson Bay Lowlands, Manitoba. Splendid examples of ice-wedge polygons, a form of patterned ground, are shown above. They occur in the permafrost peatlands of the Hudson Bay Lowlands, which are composed mainly of dry sphagnum. Brown polygons mark the location of massive ice wedges that extend from the surface down to 2 or 3 m. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Palsas are low permafrost mounds with cores of layered ***segregated ice*** and ***peat*** (**Figure 11q-8**). They are normally 1 to 7 meters high, 10 to 30 meters wide, and 15 to 150 meters long. They are also most common at the southern margin of the discontinuous permafrost zone. Palsas are believed to form when areas of reduced snow cover allow frost to penetrate more deeply into an unfrozen peat bog. This frost freezes the water in the peat forming an initial ice layer. The layer then grows in size over time as water migrates under some type of pressure from unfrozen parts of the peat to the surface of the growing ice mass.

{PRIVATE**}Figure 11q-8:** Emerging palsa in a fen bog near Churchill, Manitoba. This palsa is in the initial stages of development and is growing by absorbing some of the water in the surrounding wetland. As segregated ice builds up inside this feature, the water-loving vegetation growing on the peat is pushed up above the water table. This action causes the vegetation to die off, resulting in the bare brown peat surface on top of the mound. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

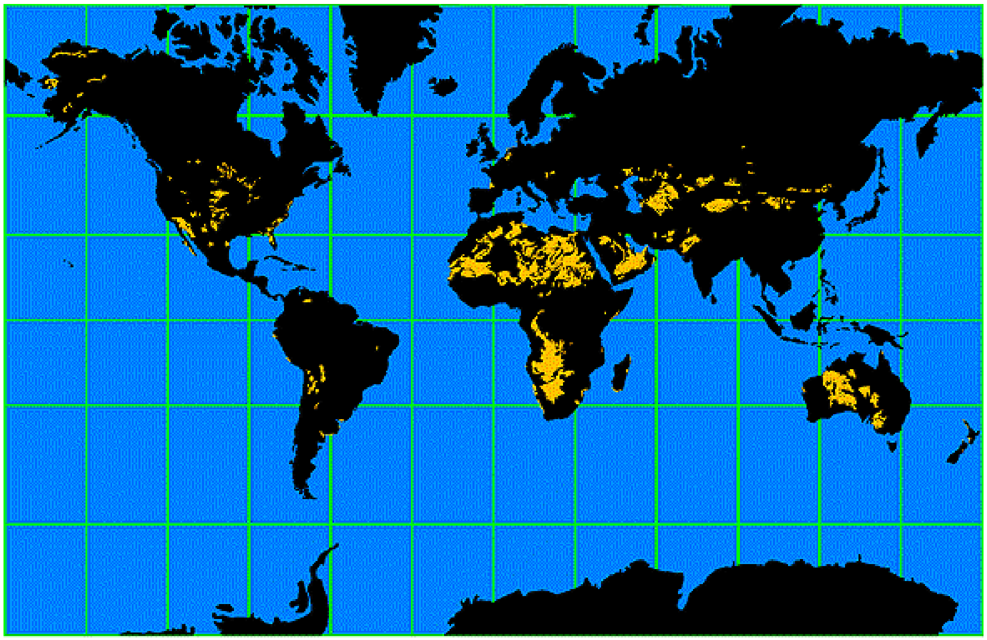
Pingos are ice-cored hills with a height between 3 to 70 meters and a diameter between 30 to 1000 meters (**Figure 11q-9**). Most pingos are circular in shape. Smaller pingos tend to have a curved top. Large ones usually have exposed ice at their top and the melting of this ice often forms a crater. Sometimes the craters are filled with water forming a lake. The ice at the core of pingos is thought to accumulate because of ***cryostatic pressure*** and ***artesian*** groundwater flow. The development of a cryostatic pingo begins with a lake with no permafrost beneath it (***talik***). The lake then gradually fills in with sediment and invading permafrost isolates the remaining water in the lake's sediments. Continued inward and downward freezing of the old lake sediments generates enough pressure to move pore water upward. This pore water then begins to freeze to form a segregated mass of ice at the core of the developing pingo. Artesian pingos develop when a supply of ***groundwater*** is channeled to a particular location where it freezes just below the ground surface.

{PRIVATE**}Figure 11q-9:** Pingos on Prince Patrick Island, Northwest Territories, Canada. These features are up to 100 meters across. Eventually the ice core will grow to such a size that it will rupture the sediment cover and become exposed. Scientists are unsure where all of the free water is coming from and how it is conducted through the frozen sediments and rock to make these features grow. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Some pingos are still actively growing. The maximum rates of vertical growth of young pingos can be as high as 1.5 meters per year. The dating of pingos has reveled that these features are generally less than 10,000 years old. Many small ones in the Arctic have ages that are less than a few hundred years old. Scientists estimate that several thousand exist in the periglacial regions of the Northern Hemisphere.

(r) Eolian Processes and Landforms

Eolian landforms are found in regions of the Earth where ***erosion*** and ***deposition*** by wind are the dominant geomorphic forces shaping the face of the landscape. Regions influenced by wind include most of the dry climates of the Earth (**Figure 11r-1**). According to the **Köppen Climate Classification System** (see **Topic *7v***), this would include regions of the world that are classified as arid deserts (**BW**) and semiarid steppe (**BS**). Wind can also cause erosion and deposition in environments where sediments have been recently deposited or disturbed. Such environments include lake and ocean coastline ***beaches***, ***alluvial fans***, and farmland where topsoil has been disturbed by cultivation.



{PRIVATE**}Figure 11r-1:** Global distribution of major deposits of eolian derived sediments.

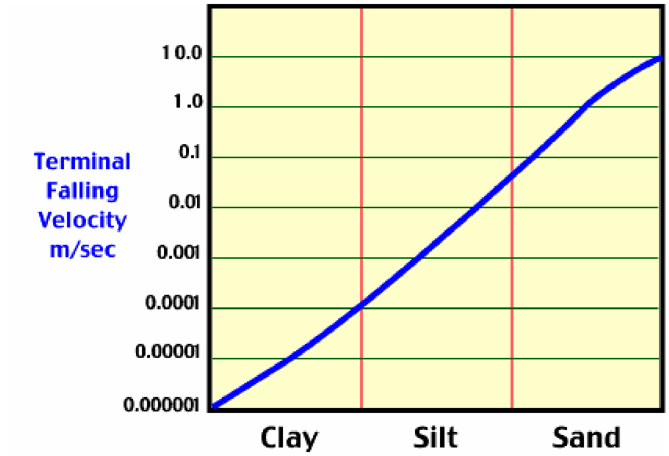
Unlike streams, wind has the ability to transport sediment up-slope as well as down-slope. The relative ability of wind to erode materials is is slight when compared to the other major erosional mediums, water and ice. Ice and water can have greater erosive power primarily because of their greater ***density***. Water is about 800 times more dense than air (density of air is 1.29 kg m^3, while the density of water is 1000 kg m^3). This physical difference limits the size of particles wind can move. The power of wind to erode surface particles is controlled primarily by two factors: wind velocity and surface roughness. Erosive force increases exponentially with increases in wind ***velocity***. For example, a velocity increase from 2 to 4 meters per second causes an eight-fold increase in erosive capacity, while an increase in wind speed from 2 to 10 meters per second generates a 125-fold increase in erosional force. Consequently, fast winds are capable of causing much more erosion than slow winds.

At ground level, the roughness of the surface plays an important role in controlling the nature of wind erosion. Boulders, trees, buildings, shrubs, and even small plants like grass and herbs can increase the frictional roughness of the surface and reduce wind velocity. Vegetation can also reduce the erosional effects of wind by binding soil particles to roots. Thus, as a general rule, the areas that show considerable amounts of wind erosion are open locations with little or no plant cover.

Threshold and Terminal Fall Velocities

Threshold velocity can be defined as velocity required to ***entrain*** a particle of a particlular size. In general, the larger the particle, the higher the threshold velocity required to move it. This law can sometimes be broken when ***clay*** sized particles are involved in the entraiment process. ***Clay*** particles have a general tendency to become cohesively bonded to each other. This aggregation results in the clumping of several particles into a mass of much larger size. As a result, the threshold velocity required to entrain clay is as great as the wind speed required to move grains of ***sand***. ***Silt*** is usually the easiest type of particle to be entrain by wind.

Terminal fall velocity can be defined as velocity at which a particle being ***transported*** by wind or water falls out and is deposited on the ground surface. **Figure 11r-2** describes the terminal fall velocities for clay, silt, and sand sized particles for wind. On this semi-log graph, a simple, some what linear, relationship is observed. The larger the particle the greater the wind speed that is required to keep it moving above the ground surface.



{PRIVATE**}Figure 11r-2:** Falling velocities for clay, silt, and sand sized particles for wind. Note the fall velocity for clay is many orders of magnitude less than the fall velocity for sand.

Sand Transport

Most of our present knowledge about the wind's ability to erode and transportation sand comes from the 1920s to 40s work of Ralph Bagnold in the deserts of North Africa. In his numerous observations and experiments dealing with sand movement, Bagnold discovered many of the key principles controlling the erosion and transport of sand in deserts.

Three different processes are responsible for the transport of sediment by wind. Wind erosion of surface particles begins when air velocities reach about 4.5 meters per second. A rolling motion called ***traction*** or ***creep*** (the later term should not to be confused with soil creep) characterizes this first movement of particles. In strong winds, particles as large as small pebbles can move through traction. About 20 to 25 percent of wind erosion is by traction. The second type of wind sediment transport involves particles being lifted off the ground, becoming suspended in the air, and then returning to the ground surface several centimeters downwind. This type of transport is called ***saltation***, and this process accounts for 75 to 80 percent of the sediment transport in dry land environments. Saltating particles are also responsible for sending additional sediment into transport. When a falling particle strikes the ground surface, part of its force of impact is transferred to another particle causing it to become airborne. Small sized particles like silt and clay have the ability to be lifted well above the zone of saltation during very strong winds and can be carried in ***suspension*** thousands of meters into the air and hundreds of kilometers downwind.

Erosional Landforms

When the force of wind is concentrated on a particular spot in the landscape, erosion can carve out a pit known as a ***deflation hollow***. Deflation hollows range in size from a few meters to a hundred meters in diameter, and may develop over several days or a couple of seasons. Much larger depressions are also found in the arid regions throughout the world. These broad, shallow depressions, called ***pans***, can cover thousands of square kilometers. One of the largest pans, known as the Qattara Depression is found in the Lybian Desert of Egypt. This pan covers around 15,000 square kilometers.

In some dry climate areas, persistent winds erode all sediments the size of ***sand*** and smaller leaving ***pebbles*** and larger particles on the ground surface. Surfaces loaded with such particles are called ***desert pavement*** or ***reg*** and sometimes resemble a worn, polished cobblestone street surface.

Depositional Landforms

Sand dunes are the most noticeable landforms produced by wind erosion and deposition of sediment. The largest ***dune fields*** are found in the Middle East and North Africa. Most large dune fields act more or less as closed systems. Once sand enters these systems, it does not leave. However, dune fields do shift across the Earth's surface from time to time. Periodic migrations of dune fields are normally caused by seasonal changes in wind direction. Over longer periods of time, dune fields may expand or contract because of climatic change. In the last few decades scientists have noticed a spatial expansion of deserts that may be correlated to human disturbance of natural vegetation cover because of agriculture.

Sand Dune Formation

Sand dunes form in environments that favor the ***deposition*** of ***sand*** (**Figure 11r-3**). Deposition occurs in areas where a pocket of slower moving air forms next to much faster moving air. Such pockets typically form behind obstacles like the ***leeward*** sides of slopes. As the fast air slides over the calm zone, saltating grains fall out of the air stream and accumulate on the ground surface.

{PRIVATE**}Figure 11r-3:** ***Wind ripples*** on top of much larger sand dunes. Wind ripples are mini-dunes between 5 centimeters and 2 meters in height and 0.1 to 5 centimeters in height. They are created by ***saltation*** when the sand grains are of similar size and wind speed is consistent. A series of wind ripples is initiate by a single irregularity in the ground surface. This irregularity launches the grains in the air and the consistences of size and windspeed cause saltation at repeated regular intervals downwind.

Dunes first begin their life as a stationary pile of sand that forms behind some type of vertical obstacle. However, when they reach a certain size threshold continued growth may also be associated with active surface migration. In a migrating dune, grains of sand are transported by wind from the ***windward*** to the ***leeward*** side and begin accumulating just over the crest. When the upper leeward slope reaches an angle of about 30-34 degrees the accumulating pile becomes unstable, and small avalanches begin to occur, moving sand to the lower part of the leeward slope. As a result of this process, the dune migrates over the ground as sand is eroded from one side and deposited on the other. This process also causes the appearance of the dune to take on a wave shape. Active movement of sand particles across the dune causes windward slope to become shallow, while the leeward slope maintains a steep slope or ***slip-face***.

{PRIVATE**}**Sand dunes begin when a ground obstruction influences the areodynamic movement of wind and ***saltating*** sand particles. As the wind blows over the obstruction, velocity is reduced causing particles to fall out of the air stream. The pile of particles builds both vertically and horizontally, and the dune is at first stationary. The dune continues to grow vertically, and at a certain height the top of the feature begins to encounter faster moving air. This causes particles to be lifted from the ***windward*** side of dune. As these lifted particles move forward they reach areas of slower wind speeds at the ***leeward*** end of the dune, causing the particles to fall out of the air. The net result of this process is the migration of sand particles from the front of the dune to its backside. It also causes the whole dune to move forward.

The velocity of the wind above the ground surface determines the height of a dune. The maximum height is variable but usually falls in the range of 10 to 25 meters. In most cases, dune height is a function of surface friction. Height growth stops when friction can no longer slow the wind flowing over the dune to a point where deposition occurs. The tallest sand dunes in the world are found in Saudi Arabia and measure more than 200 meters. However, these features are not individual dunes, but a massive complex of sand dunes that forms when smaller, faster moving dunes migrate onto larger, slower moving dunes.

Desert Dunes

Desert sand dunes occur in an amazing diversity of forms. **Table 11r-1**describes the major types of dunes classified by geomo rphologists.

Table 11r-1: Major types of sand dunes.

{PRIVATE**}Type**	**Description**
Barchan	Crescent-shaped dune whose long axis is transverse to the dominant wind direction. The points of the dune, called the wings of the barchan, are curved downwind and partially enclosing the slip-face. Barchans usually form where there is a limited supply of sand, reasonably flat ground, and a fairly even flow of wind from one direction. Single slipface.
Transverse	Long asymmetrical dunes that form at right angles to the wind direction. Form when there is an abundant supply of sand and relatively weak winds. These dunes have a single long slip-face.
Parabolic	Crescent-shaped dune whose long axis is transverse to the dominant wind direction. The points of this dune curve upwind. Multiple slip-faces. These dunes form when scattered vegetation stabilizes sediments and a U-shaped blowout forms between clumps of plants.
Barchanoid Ridge	Is a long, asymmetrical dune that runs at right angles to the prevailing wind direction. A barchanoid ridge consists of several joined barchan dunes and looks like a row of connected crescents. Each of the barchan dunes produces a wave in the barchanoid ridge. Occurs when sand supply is greater than in the conditions that create a barchan dune.
Longitudinal	Sinuous dune that can be more than 100 kilometers long and 100 meters high. Created when there are strong winds from at least two directions. The dune ridge is symetrical, aligned parallel to the net direction of the wind, and has slipfaces on either side. See **Figure 4r-5** below.
Seif	Sub-type of longitudinal dune that is shorter and has a more sinuous ridge.
Star Dune	Large pyramidal or star-shaped dune with three or more sinuous radiating ridges from a central peak of sand. This dune has 3 or more slipfaces. Produced by variable winds. This dune does not migrate along the ground, but does grow vertically.
Dome	Mound of sand that is circular or elliptical in shape. Has no slip-faces. May be formed by the modification of stationary barchans.
Reversing	Dune that is intermediate between a star and transverse dunes. Ridge is asymmetrical and has two slip-faces.

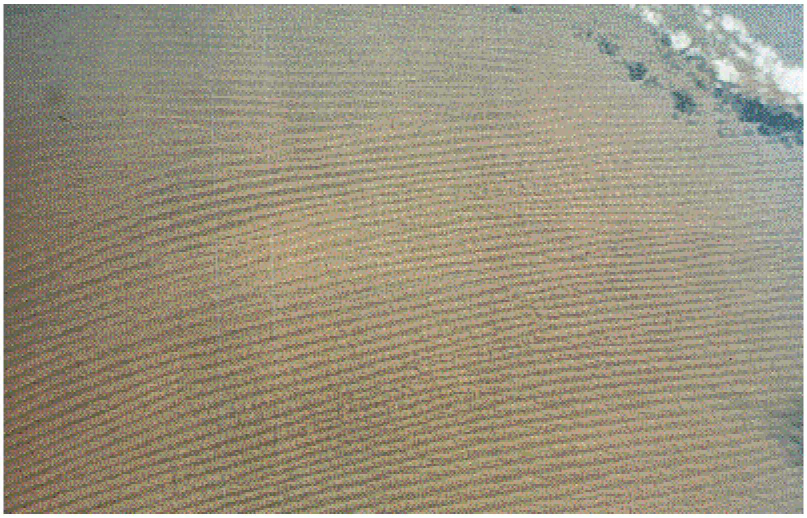


Figure 11r-5: Longitudinal dunes, Arabian Peninsula (**Source:** *NASA*).

Coastal Dunes

Active sand dune formation is also found on the coasts of the continents. Coastal dunes form when there is a large supply of beach sand and strong winds blowing from sea to shore (**Figure 11r-5**). The beach area must also be wide and sufficiently influenced by wave action to keep it free of plants.

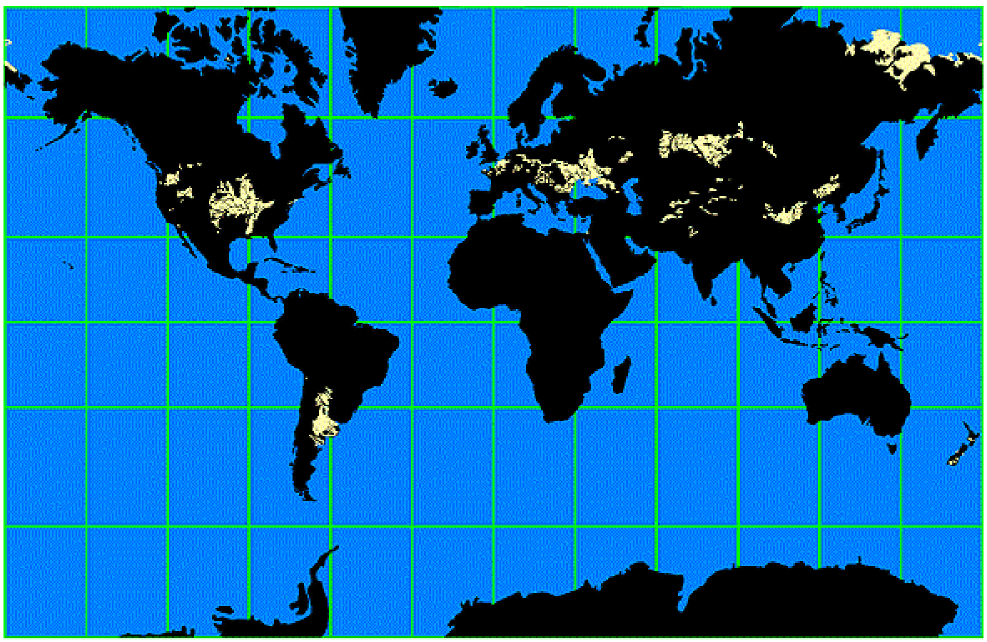
Many coastal dune deposits develop in association with ***blowouts*** in ridges of beach sand. Blowouts are small saucer shaped depressions where there is a deposit of sand at the upwind end of the feature. As wind erosion continues, the deposit grows and begins to migrate inland forming a parabolic sand dune. The flanks of these dunes tend to be more stable and are often colonized by plants like dune grass, sea oats, and sand cherry. This colonization by plants re-inforces the stability of the dune's flanks.

Coastal dunes are dissimilar from desert dunes in their form and shape and the fact that they do not migrate. The presence of vegetation limits migration and radically modifies the dune environment by altering the patterns of airflow, reducing sand erosion, and stabilizing the leeslope of the dune.

{PRIVATE**}Figure 11r-5:** Vegetated beach dune. (**Source:** *Natural Resources Canada - Terrain Sciences Division - Canadian Landscapes*).

Loess Deposits

Loess is another major deposit created by wind. Less visible than sand dunes, loess is found over large areas of the Earth (**Figure 11r-6**). It is also important for humans because it creates very fertile soils. Large deposits of loess exist in northeastern China, central plains of the United States, Pampas of Argentina, the Ukraine, and central Europe. Loess is mainly composed of ***silt***. Because of its small size it can be held in suspension and carried great distances by wind. Most loess deposits appear to have been formed by winds that blew over ***glacial*** deposits during the ***Pleistocene***. The major deserts of the world also appear to have produced significant amounts of loess. Recent research has uncovered that soils in the Amazon basin may have been enriched with loess deposits that originated from African deserts.



{PRIVATE**}Figure 11r-6:** Global distribution of major deposits of loess sediments.

Bibliografía

Capítulo 1. Introduction to Physical Geography

Gregory, K.J. 1985. The Nature of Physical Geography. Edward Arnold, London.
Holt-Jensen, A. 1988. Geography: Its Histroy and Concepts. Second Edition. Barnes and Noble Books, Totowa, New Jersey.
Lanegran, D.A. and R. Palm (editors). 1978. Invitation to Geography. Second Edition. McGraw-Hill Publishing, New York.
Livingston, D.N. 1992. The Geographical Tradition. Basil Blackwell, Cambridge, Mass.
Lukermann, F. 1964. Geography as a formal intellectual discipline and the way in which it contributes to human knowledge. The Canadian Geographer 8(4): 167-172.
Marshall, G.J. and P.E. James. 1993. All Possible Worlds: A History of Geographical Ideas. John Wiley and Sons, New York.
Morrill, R.L. 1983. The nature, unity and value of Geography. Professional Geographer 35(1): 1-9.
Minshull, Roger M. 1970. The Changing Nature of Geography. Hutchinson, London.
Pattison, W.D. 1964. The four traditions of Geography. Journal of Geography 63: 211-216.
Robinson, J.L. 1976. A new look at the four traditions of Geography. Journal of Geography 75: 520-530.

Capítulo 2. Maps, Remote Sensing, and GIS

Campell, J.B. 1991. Introductory Cartography. Second Edition. Wm. C. Brown Publishers, Dubuque, IA.
Campell, J.B. 1996. Introduction to Remote Sensing. Second Edition. The Guilford Press, New York.
Chrisman, N. 1997. Exploring Geographic Information Systems. John Wiley and Sons, New York.
Clarke, K.C. 2001. Getting Started with Geographical Information Systems. Third Edition. Prentice Hall, Upper Saddle River, New Jersey.
DeMers, M.N. 2000. Fundamentals of Geographic Information Systems. Second Edition. John Wiley and Sons, New York.
Dent, B.D. 1999. Cartography: Thematic Map Design. Fifth Edition. WCB/McGraw-Hill, Dubuque, IA.
Heywood, I., S. Cornelius, and S. Carver. 1998. An Introduction to Geographical Information Systems. Prentice Hall, Upper Saddle River, New Jersey.
Holt-Jensen, A. 1988. Geography: Its History and Concepts. Second Edition. Barnes and Noble, Totowa, New Jersey.
Maguire, D.J. 1989. Computers in Geography. John Wiley and Sons, New York.
Mather, P.M. 1991. Computer Applications in Geography. John Wiley and Sons, New York.
Rees, W.G. 1990. Physical Principles of Remote Sensing. Cambridge University Press, Cambridge.
Robinson, A.H., J.L. Morrison, P.C. Muehrcke, A.J. Kimering, and S.C. Guptill. 1995. Elements of Cartography. Sixth Edition. John Wiley and Sons, New York.
Sabins, F.J. Jr. 1996. Remote Sensing: Principles and Interpretation. Third Edition. W.H. Freeman and Company, New York.
Star, J., and J. Estes. 1990. Geographic Information Systems. Prentice Hall, Upper Saddle River, New Jersey.
Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, Harmondsworth.

Capítulo 3. The Science of Physical Geography

Barber, G.M. 1988. Elementary Statistics for Geographers. The Guilford Press, New York.
Clark, W.A.V. and P.L. Hoskins. 1986. Statistical Methods for Geographers. John Wiley, New York.
Ebdon, D. 1985. Statistics in Geography. Second Edition. Basil Blackwell, Oxford.
Gregory, K.J. 1985. The Nature of Physical Geography. Edward Arnold, London.
Griffith, D.A. and C.G. Amrhein. 1991. Statistical Analysis for Geographers. Prentice Hall, Englewood Cliffs, New Jersey.
Haring, L.L., J.F. Lounsbury, and J.W. Frasier. 1992. Introduction to Scientific Geographic Research. Fourth Edition. Wm. C. Brown, Dubuque, Iowa.
Hecht, J. 1995. The geological timescale. Inside Science 81. New Scientist May 20.

Holton, G., H. Chang, and E. Jurkowitz. 1996. How a scientific discovery is made: a case history. American Scientist 84: 364-376.

Matthews, R. 1997. Beyond space and time. New Scientist May 17: 38-42.

McGrew, J.C. and C.B. Monroe. 1993. An Introduction to Statistical Problem Solving in Geography. Wm. C. Brown, Dubuque, Iowa.

Shaw, G. abd D. Wheeler. 1994. Statistical Techniques in Geographical Analysis. Second Edition. David Fulton Publishers, London.

Silk, J. 1979. Statistical Concepts in Geography. George Allen and Unwin, Boston.

Walford, N. 1994. Geographic Data Analysis. John Wiley and Sons, New York.

Capítulo 4. Introduction to Systems Theory

Odum, H.T. 1983. Systems Ecology: An Introduction. John Wiley and Sons, New York.

Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, London.

Capítulo 5. The Universe, Earth, Natural Spheres, and Gaia

Allegre, C.J., and S.H. Schneider 1994. The evolution of the Earth. Scientific American (Oct.): 66-75.

Bucher, M.A. and D.N. Spergel. 1999. Inflation in a low-density Universe. Scientific American (Jan.): 62-69.

Chown, M. 1994. Birth of the universe. Inside Science 69. New Scientist Feb. 26.

Croswell, K. 1997. Cosmic cannibals. New Scientist July 12: 28-32.

Davies, P. 1996. The day time began. New Scientist April 27: 30-35.

Henbest, N. 1990. Rocky dwarfs and gassy giants. Inside Science 30. New Scientist Feb. 10.

Joseph, Lawrence E. 1990. Gaia : The Growth of an Idea. St. Martin's Press, New York.

Hogan, CJ 1996. Primordial deuterium and the big bang. Scientific American (Dec.): 68-73.

Hogan, CJ, R.P. Kirshner, and N.B. Suntzeff. 1999. Surveying space-time with supernovae. Scientific American (Jan.): 46-51.

Kaku, M. 1997. Into the eleventh dimension. New Scientist Jan 18: 32-36.

Kasting, J.F. 1993. Earth's early atmosphere. Science 259: 920-926.

Krauss, L.M. 1999. Cosmological antigravity. Scientific American (Jan.): 53-59.

Lovelock, J.E. 1988. The Ages of Gaia : A Biography of Our Living Earth. Norton, New York.

Lovelock, J.E. 1979. Gaia, a New Look at Life on Earth. Oxford University Press, Oxford.

Macchetto, F.D., and M. Dickinson. 1997. Galaxies in the young universe. Scientific American (May): 92-99.

Margulis, Lynn, and L. Olendzenski. 1992. Environmental Evolution : Effects of the Origin and Evolution of Life on Planet Earth. Cambridge University Press, Cambridge.

Matthews, R. 1997. Beyond space and time. New Scientist May 17: 38-42.

Orgel, L.E. 1994. The origin of life on the Earth. Scientific American (Oct.): 77-83.

Peebles, P.J.E., D.N. Schramm, E.L. Turner, and R.G. Kron. 1994. The evolution of the universe. Scientific American (Oct.): 53-57.

Press, F., and R. Siever. 1986. Earth. W.H. Freeman, New York.

Rambler, M. B., L. Margulis and F. Reni. 1988. Global Ecology : Towards a Science of the Biosphere. Academic Press, Boston.

Van Andel, J. H. 1994. New Views on an Old Planet: A History of Global Change. Cambridge University Press, Cambridge.

Walker, G. 1997. God's firecrackers. New Scientist May 31: 28-32.

Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, London.

Capítulo 6. Energy and Matter

Gribbin, J. 1993. Inside the sun. Inside Science 60. New Scientist March 13.

Foukal, P.V. 1990. The variable sun. American Scientist (Feb.): 34-41.

Kaku, M. 1997. Into the eleventh dimension. New Scientist Jan 18: 32-36.

Lang, K.R. 1997. SOHO reveals the secrets of the sun. American Scientist (March): 40-47.

Nesme-Ribes, E., S.L. Baliunas, and D. Sokoloff. 1996. The stellar dynamo. American Scientist (Aug.): 46-52.

Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, London.

Capítulo 7. Introduction to Meteorology and Climatology

Ahrens, C.D. 1994. Meteorology Today: An Introduction to Weather, Climate, and the Environment. Fifth Edition. West Pub. Co., St. Paul.
Allegre, C.J., and S.H. Schneider 1994. The evolution of the earth. Scientific American (Oct.): 66-75.
Barry, R.G., and R.J. Chorley. 1987. Atmosphere, Weather, and Climate. Methuen, London.
Budyko, M.I. 1974. Climate and Life. Academic Press, New York.
Eagleman, J.R. 1985. Meteorology: The Atmosphere in Action. Wadsworth, Belmont, California.
Gribbin, J. 1988. The greenhouse effect. Inside Science 13. New Scientist Oct. 22.
Gribbin, J. 1988. The ozone layer. Inside Science 9. New Scientist May 5.
Henderson-Sellers, A., and P.J. Robinson. 1986. Contemporary Climatology. Longman Scientific and Technical, London.
Kasting, J.F. 1993. Earth's early atmosphere. Science 259: 920-926.
Lockwood, J.G. 1979. Causes of Climate. Wiley, New York.
Lydolph, P.E. 1985. Weather and Climate. Rowman and Allanheld, Totowa, New Jersey.
Minnis, P., E.F. Harrison, L.L. Stowe, G.G. Gibson, F.M. Denn, D.R. Doelling, and W.L. Smith, Jr. 1993. Radiative climate forcing by the Mt. Pinatubo eruption. Science 259: 1369-1508.
Moran, J.M., and M.D. Morgan. 1997. Meteorology: The Atmosphere and the Science of Weather. Fifth Edition. Prentice Hall, Upper Saddle River, New Jersey.
Musk, L.F. 1988. Weather Systems. Cambridge University Press, Cambridge.
O'Hare, G. 1986. The Atmosphere System: An Introduction to Meteorology and Climatology. Oliver and Boyd, Edinburgh.
Oke, T.R. 1990. Boundary Layer Climates. Second Edition. Routledge, London.
Oliver, J.E., and R.W. Fairbridge. 1987. The Encyclopedia of Climatology. Van Nostrand Reinhold, New York.
Parker, S.P. 1988. Meteorology Source Book. McGraw-Hill, New York.
Pearce, F. 1997. Greenhouse wars. New Scientist July 19: 38-43.
Shelton, M.L. 1988. Climate and Weather: A Spatial Perspective. Kendall/Hunt, Dubuque, Iowa.
Sumner, G.N. 1988. Precipitation: Process and Analysis. Wiley, New York.
Taylor, K. 1999. Rapid climate change. American Scientist 87: 320-327.
Tett, S.F.B., P.A. Stott, M.R. Allen, W.J. Ingram, and J.F.B. Michell. 1999. Causes of twentieth-century temperature change near the Earth's surface. Nature 399: 569-572.
Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, London.
Wright, D. 1983. Meteorology. Blackwell, Oxford.

Capítulo 8. Introduction to Hydrology

Bras, R.L. 1989. Hydrology: An Introduction to Hydrologic Science. Addison-Wesley, Reading, Mass.
Jones, P. 1983. Hydrology. Blackwell, Oxford.
Matthews, R. 1997. Wacky water. New Scientist June 21: 40-43.
Morisawa, M., and K.M. Clayton. 1985. Rivers: Form and Process. Longman, London.
Pearce, F. 1987. Acid rain. Inside Science. New Scientist Nov. 5.
Price, M. 1991. Water from the ground. Inside Science 42. New Scientist Feb. 16.
Raghunath, H.M. 1985. Hydrology: Principles, Analysis, and Design. Wiley, New York.
Sumner, G.N. 1988. Precipitation: Process and Analysis. Wiley, New York.
Viessman, W., G.L. Lewis, and J.W. Knapp. 1989. Introduction to Hydrology. Harper and Row, New York.
Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, London.

Capítulo 9. Introduction to Biogeography and Ecology

Barbour, M.G., J.H. Banks, and W.D. Pitts. 1987. Terrestrial Plant Ecology. Benjamin/Cummings, Menlo Park, California.

Begon, M., J.L. Harper, and C.R. Townsend. 1986. Ecology: Individuals, Populations and Communities. Sinauer Associates, Sunderland, Mass.
Cohen, P. 1996. Let there be life. New Scientist 151(2037): 22-27.
Colinaux, P.A. 1986. Ecology. Wiley, New York.
Cox, C.P., I.N. Healey, and P.D. Moore. 1976. Biogeography: An Ecological and Evolutionary Approach. Blackwell, London.
Ehrlich, P.R., and J. Roughgarden. 1987. The Science of Ecology. MacMillan, New York.
Gould, S.J. 1994. The evolution of life on the Earth. Scientific American (Oct.): 84-91.
Krebs, C.J. 1978. Ecology: The Experimental Analysis of Distribution and Abundance. Harper and Row, New York.
MacArthur, R.H. 1984. Geographical Ecology: Patterns in the Distribution of Species. Princeton University Press, Princeton, New Jersey.
Mielke, H.W. 1989. Patterns of Life: Biogeography of a Changing World. Unwin Hyman, Boston.
Myers, A.A., and P.S. Giller. 1988. Analytical Biogeography: An Integrated Approach to the Study of Animal and Plant Distributions. Chapman and Hall, London.
Odum, E.P. 1993. Ecology and Our Endangered Life-Support Systems. Sinauer Associates, Sunderland, Mass.
Odum, H.T. 1983. Systems Ecology: An Introduction. Wiley, New York.
Orgel, L.E. 1994. The origin of life on the Earth. Scientific American (Oct.): 77-83.
Pears, N. 1977. Basic Biogeography. Longman, London.
Pennisi, E., and W. Roush. 1997. Developing a new view of evolution. Science 277: 34-37.
Simmons, I.G. 1982. Biogeographical Processes. Allen and Unwin, London.
Smith, R.L. 1990. Ecology and Field Biology. Haper Collins, New York.
Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, Harmondsworth.

Capítulo 10. Introduction to Geology

Allegre, C.J., and S.H. Schneider 1994. The evolution of the Earth. Scientific American (Oct.): 66-75.
Bowler, S. 1993. Earthquakes and volcanoes. Inside Science 64. New Scientist Sept. 18.
Bridges, E.M. 1990. World Geomorphology. Cambridge University Press, Cambridge.
Dolgoff, A. 1998. Physical Geology. Houghton Mifflin Company, Boston.
Easterbrook, D.J.1993. Surface Processes and Landforms. MacMillan Publishing Company, New York.
Fifield, R. 1988. Structure of the Earth. Inside Science. New Scientist Feb. 25.
Gerrard, J. 1988. Rocks and Landforms. Unwin Hyman, London.
Hecht, J. 1995. The geological timescale. Inside Science 81. New Scientist May 20.
Monroe, J.S., and R. Wicander. 2001. Physical Geology: Exploring the Earth. 4th Edition. Brooks?Cole, Pacific Grove, California.
Ollier, C., and C. Pain. 2000. The Origin of Mountains. Routledge, London.
Plummer, C.C., D. McGeary, and D.H. Carlson. 1999. Physical Geology. 8th Edition. WCB McGraw-Hill, Dubuque, Iowa.
Press, F., and R. Siever. 1986. Earth. W.H. Freeman, New York.
Thompson, G.R. and J. Turk. 1991. Modern Physical Geology. Saunders, Toronto.
van Andel, J. H. 1994. New Views on an Old Planet: A History of Global Change. Cambridge University Press, Cambridge.
Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, Harmondsworth.

Capítulo 11. Introduction to Geomorphology

Bloom, A.L. 1998. Geomorphology: A Systematic Analysis of Late Cenozoic Landforms. Third Edition. Pretice Hall, Upper Saddle River, New Jersey.
Bridges, E.M. 1990. World Geomorphology. Cambridge University Press, Cambridge.
Easterbrook, D.J.1993. Surface Processes and Landforms. MacMillan Publishing Company, New York.
Gerrard, J. 1988. Rocks and Landforms. Unwin Hyman, London.
Ollier, C., and K.M. Clayton. 1984. Weathering. Longman, London.
Press, F., and R. Siever. 1986. Earth. W.H. Freeman, New York.

Selby, M.J. 1985. Earth's Changing Surface: An Introduction to Geomorphology. Oxford University Press, Oxford.

Thompson, G.R. and J. Turk. 1991. Modern Physical Geology. Saunders, Toronto.

Trenhaile, A.S. 1990. The Geomorphology of Canada: An Introduction. Oxford University Press, New York.

van Andel, J. H. 1994. New Views on an Old Planet: A History of Global Change. Cambridge University Press, Cambridge.

Whitlow, J.B. 1984. The Penguin Dictionary of Physical Geography. Penguin, Harmondsworth.

Related Internet Links

Academic Listservs, Forums, Organizations, and Societies
American Geographers Association (http://www.cla.sc.edu/gis/aaggis/index.html)
Association of American Geographers (http://www.aag.org)
Association of Polish Geomorphologists (http://main.amu.edu.pl/~sgp/wel.htm)
Canadian Association of Geographers - Western Division (http://office.geog.uvic.ca/dept/wcag/wcag.htm)
Canadian Association of Geographers (http://zeus.uwindsor.ca/cag/cagindex.html)
CTI Centre for Geography, Geology, and Meteorology (http://www.geog.le.ac.uk/cti/)
Dutch Geographical Society (http://www.knag.nl/)
Economic Geography Research Group (http://www.econgeog.org.uk/)
Geographical Society of Finland (http://www.helsinki.fi/ml/maant/geofi/)
Geographical Society of Ireland (http://www.ucd.ie/~geograph/)
Geography Education Specialty Group (http://www.colorado.edu/geography/COGA/geoed/)
Historical Geography Speciality Group (http://www.geog.okstate.edu/hgsg/hgsg.htm)
International Glaciological Society (http://www.geog.leeds.ac.uk/groups/igs/igshome.htm)
Latin Americanist Geographers (http://clagonline.org/)
National Center for Geographic Information & Analysis (http://www.ncgia.ucsb.edu/ncgia.html)
National Council for Geographic Education (http://www2.oneonta.edu/~baumanpr/ncge/rstf.htm)
National Geographic Society (http://www.nationalgeographic.com/)
New Zealand Geographical Society (http://www.waikato.ac.nz/wfass/subjects/geography/nzgs/nzgs.htm)
Polish Academy of Sciences Institute of Geography (http://www.igipz.pan.pl/)
Political Geography Speciality Group (http://garnet.acns.fsu.edu/~dpurcell/pgsg1.html)
Population Specialty Group (http://www.pop.psu.edu/aag/psg.html)
Royal Geographical Society of Queensland (http://www.rgsq.gil.com.au/)
Royal Geographical Society of South Australia (http://www.archivists.org.au/directory/data/301.htm)
Royal Geographical Society (http://www.rgs.org/)
Royal Scottish Geographical Society (http://www.geo.ed.ac.uk/home/scotland/rsgs/)
The Geographical Association (http://www.geography.org.uk/)
University Texas (Austin): Newsgroups and Listservers
(http://www.utexas.edu/depts/grg/virtdept/resources/listserv/listserv.htm)
Definitions of Geography
Various Definitions of Geography (http://www.westga.edu/~geosci/)
History of Geography
Age of Exploration Timeline (http://www.mariner.org/age/histexp.html)
Encarta Online (http://encarta.msn.com/encnet/features/reference.aspx)
Ptolemy's Geography - Science of the Earth's Surface (http://www.ibiblio.org/expo/vatican.exhibit/exhibit/d-mathematics/Ptolemy_geo.html)
Journal and Publications
Applied Geography (http://www.elsevier.nl/inca/publications/store/3/0/3/9/0/)
Canadian Geographic (http://www.cangeo.ca/)
Cartography - The Journal (http://www.mappingsciences.org.au/journal.htm)
CIA Maps and Publications (http://www.odci.gov/cia/publications/pubs.html)
Current Geographical Publications (http://leardo.lib.uwm.edu/cgp/)
Earth Surface Processes and Landforms (http://www.interscience.wiley.com/jpages/0197-9337/)
Geografiska Annaler Series A: Physical Geography (http://www.blackwellpublishing.com/journal.asp?ref=0435-3676)
Geografiska Annaler Series B: Human Geography (http://www.blackwellpublishing.com/journal.asp?ref=0435-3684)
Geographical Abstracts: Physical Geography (http://www.elsevier.nl/inca/publications/store/4/0/5/8/8/6/)
GeoJournal (http://www.kluweronline.com/issn/0343-2521)
Journal of Cultural Geography (http://www.geog.okstate.edu/users/culture/culture.htm)
Journal of Political Ecology (http://dizzy.library.arizona.edu/ej/jpe/jpeweb.html)
Mapping Sciences and Remote Sensing (http://www.bellpub.com/msrs/index.html)
Nature (http://www.nature.com/)
New Scientist (http://www.newscientist.com/)
New Zealand Geographer (http://www.waikato.ac.nz/wfass/subjects/geography/nzgs/geographer.htm)
Photogrammetric Engineering and Remote Sensing (http://www.asprs.org/asprs/publications/pe%26rs/index.html)
Science (http://science-mag.aaas.org/)
Scientific American (http://www.sciam.com/)
Singapore Journal of Tropical Geography (http://www.blackwellpublishing.com/journal.asp?ref=0129-7619)

The Bulletin of the Society of Cartographers (http://www.shef.ac.uk/uni/projects/sc/bulletin/bulletin.htm)
The Cartographic Journal (http://www.cartography.org.uk/)
The Canadian Geographer / Le Géographe Canadien (http://www.geog.ouc.bc.ca/tcg/)

WWW Indexes

About.com Guide to Geography (http://geography.about.com/)
Geography Departments Worldwide (http://geowww.uibk.ac.at/geolinks/index.html)
University of Texas (Austin): Resources for Geographers
(http://www.colorado.edu/geography/virtdept/resources/contents.htm)
WWW Virtual Library - Geography (http://geography.pinetree.org/)