Part I

INTRODUCING LANDFORMS AND LANDSCAPES

1

WHAT IS GEOMORPHOLOGY?

Geomorphology is the study of landforms and the processes that create them. This chapter covers:

- historical, process, applied, and other geomorphologies
- the form of the land
- land-forming processes and geomorphic systems
- the history of landforms
- methodological isms

INTRODUCING GEOMORPHOLOGY

The word geomorphology derives from three Greek words: γεω (the Earth), μορφη (form), and λογος (discourse). Geomorphology is therefore 'a discourse on Earth forms'. It is the study of Earth's physical land-surface features, its landforms – rivers, hills, plains, beaches, sand dunes, and myriad others. Some workers include submarine landforms within the scope of geomorphology. And some would add the landforms of other terrestrial-type planets and satellites in the Solar System – Mars, the Moon, Venus, and so on. Landforms are conspicuous features of the Earth and occur everywhere. They range in size from molehills to mountains to major tectonic plates, and their 'lifespans' range from days to millennia to aeons (Figure 1.1).

Geomorphology was first used as a term to describe the morphology of the Earth's surface in the 1870s and 1880s (e.g. de Margerie 1886, 315). It was originally defined as 'the genetic study of topographic forms' (McGee 1888, 547), and was used in popular parlance by 1896. Despite the modern acquisition of its name, geomorphology is a venerable discipline (Box 1.1). It investigates landforms and the processes that fashion them. A large corpus of geomorphologists expends much sweat in researching relationships between landforms and the processes acting on them now. These are the **process** or **functional geomorphologists**. Many geomorphic processes affect,

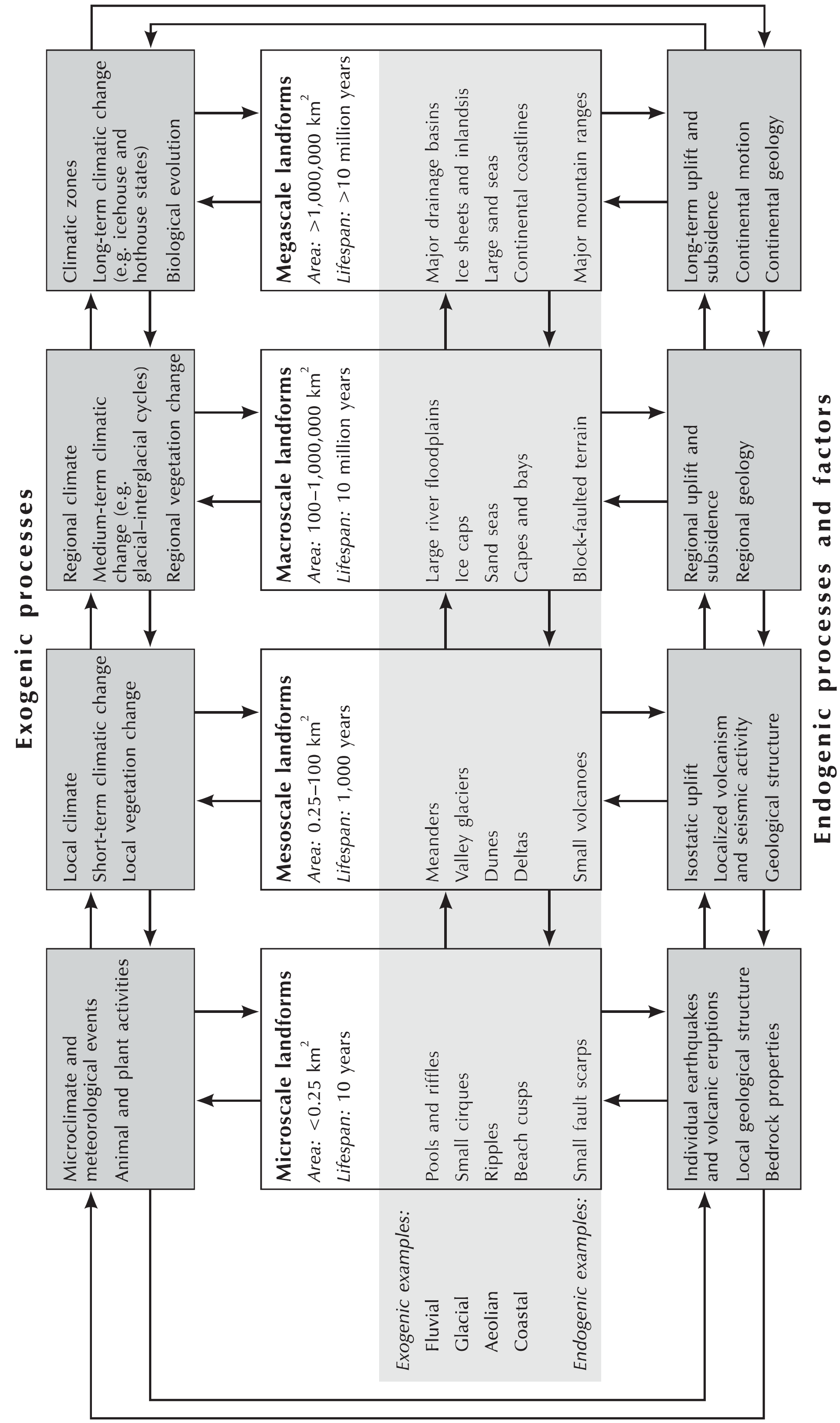


Figure 1.1 Landforms at different scales and their interactions with exogenic and endogenic processes.

Box 1.1

THE ORIGIN OF GEOMORPHOLOGY

Ancient Greek and Roman philosophers wondered how mountains and other surface features in the natural landscape had formed. Aristotle, Herodotus, Seneca, Strabo, Xenophanes, and many others discoursed on topics such as the origin of river valleys and deltas, and the presence of seashells in mountains. Xenophanes of Colophon (*c.* 580–480 BC) speculated that, as seashells are found on the tops of mountains, the surface of the Earth must have risen and fallen. Herodotus (*c.* 484–420 BC) thought that the lower part of Egypt was a former marine bay, reputedly saying 'Egypt is the gift of the river', referring to the year-by-year accumulation of river-borne silt in the Nile delta region. Aristotle (384–322 BC) conjectured that land and sea change places, with areas that are now dry land once being sea and areas that are now sea once being dry land. Strabo (64/63 BC–AD 23?) observed that the land rises and falls, and suggested that the size of a river delta depends on the nature of its catchment, the largest deltas being found where the catchment areas are large and the surface rocks within it are weak. Lucius Annaeus Seneca (4 BC–AD 65) appears to have appreciated that rivers possess the power to erode their valleys. About a millennium later, the illustrious Arab scholar ibn-Sina, also known as Avicenna (980–1037), who translated Aristotle, propounded the view that some mountains are produced by differential erosion, running water and wind hollowing out softer rocks. During the Renaissance, many scholars debated Earth history. Leonardo da Vinci (1452–1519) believed that changes in the levels of land and sea explained the presence of fossil marine shells in mountains. He also opined that valleys were cut by streams and that streams carried material from one place and deposited it elsewhere. In the eighteenth century, Giovanni Targioni-Tozzetti (1712–84) recognized evidence of stream erosion. He argued that the valleys of the Arno, Val di Chaina, and Ombrosa in Italy were excavated by rivers and floods resulting from the bursting of barrier lakes, and suggested that the irregular courses of streams relate to the differences in the rocks in which they cut, a process now called differential erosion. Jean-Étienne Guettard (1715–86) argued that streams destroy mountains and the sediment produced in the process builds floodplains before being carried to the sea. He also pointed to the efficacy of marine erosion, noting the rapid destruction of chalk cliffs in northern France by the sea, and the fact that the mountains of the Auvergne were extinct volcanoes. Horace-Bénédict de Saussure (1740–99) contended that valleys were produced by the streams that flow within them, and that glaciers may erode rocks. From these early ideas on the origin of landforms arose modern geomorphology. (See Chorley *et al.* 1964 and Kennedy 2005 for details on the development of the subject.)

and are affected by, human activities. Applied geomorphologists explore this rich area of enquiry, which is largely an extension of process geomorphology. Many landforms have a long history, and their present form does not always relate to the current processes acting upon them. The nature and rate of geomorphic processes change with time, and some landforms were produced under different environmental conditions, surviving today as relict features. In high latitudes, many landforms are relicts from the Quaternary glaciations; but, in parts of the world, some landforms survive from millions and hundreds of millions of years ago. Geomorphology, then, has an important historical dimension, which is the domain of the **historical geomorphologists**. In short, modern geomorphologists study three chief aspects of landforms – **form**, **process**, and **history**. The first two are sometimes termed functional geomorphology, the last historical geomorphology (Chorley 1978). Process studies have enjoyed hegemony for some three or four decades. Historical studies were sidelined by process studies but are making a strong comeback. Although process and historical studies dominate much modern geomorphological enquiry, particularly in English-speaking nations, other types of study exist. For example, **structural geomorphologists**, who were once a very influential group, argued that underlying geological structures are the key to understanding many landforms. **Climatic geomorphologists**, who are found mainly in France and Germany, believe that climate exerts a profound influence on landforms, each climatic region creating a distinguishing suite of landforms (p. 13).

Historical geomorphology

Traditionally, historical geomorphologists strove to work out landscape history by mapping morphological and sedimentary features. Their golden rule was the dictum that '**the present is the key to the past**'. This was a warrant to assume that the effects of geomorphic processes seen in action today may be legitimately used to infer the causes of assumed landscape changes in the past. Before reliable dating techniques were available, such studies were difficult and largely educated guesswork. However, the brilliant successes of early historical geomorphologists should not be overlooked.

William Morris Davis

The '**geographical cycle**', expounded by **William Morris Davis**, was the first modern theory of landscape evolution (e.g. Davis 1889, 1899, 1909). It assumed that uplift takes place quickly. Geomorphic processes, without further complications from tectonic movements, then gradually wear down the raw topography. Furthermore, slopes within landscapes decline through time – maximum slope angles slowly lessen (though few field studies have substantiated this claim). So topography is reduced, little by little, to an extensive flat region close to baselevel – **a peneplain** – with occasional hills, called **monadnocks** after Mount Monadnock in New Hampshire, USA, which are local erosional remnants, standing conspicuously above the general level. The reduction process creates a time sequence of landforms that progresses through the stages of **youth**, **maturity**, and **old age**. However, these terms, borrowed from biology, are misleading and much censured (e.g. Ollier 1967; Ollier and Pain 1996, 204–5). The 'geographical cycle' was designed to account for the development of humid temperate landforms produced by prolonged wearing down of uplifted rocks offering uniform resistance to erosion. It was extended to other landforms, including arid landscapes, glacial landscapes, periglacial landscapes, to landforms produced by shore processes, and to karst landscapes.

William Morris Davis's 'geographical cycle' – in which landscapes are seen to evolve through stages of youth, maturity, and old age – must be regarded as a classic work, even if it has been superseded (Figure 1.2). Its appeal seems to have lain in its theoretical tenor and in its simplicity (Chorley 1965). It had an all-pervasive influence on geomorphological thought and spawned the once highly influential field of denudation chronology. The work of denudation chronologists, who dealt mainly with morphological evidence, was subsequently criticized for seeing flat surfaces everywhere.

Walther Penck

A variation on Davis's scheme was offered by **Walther Penck**. According to the Davisian model, uplift and planation take place alternately. But, in many landscapes, uplift and denudation occur at the same time. The continuous and gradual interaction of tectonic processes and denudation leads to a different model of landscape evolution, in which the evolution of individual slopes is thought to determine the evolution of the entire landscape (Penck 1924, 1953). Three main slope forms evolve with different combinations of uplift and denudation rates. First, convex slope profiles, resulting from waxing development (*aufsteigende Entwicklung*), form when the uplift rate exceeds the denudation rate. Second, straight slopes, resulting from stationary (or steady-state) development (*gleichförmige Entwicklung*), form when uplift and denudation rates match one another. And, third, concave slopes, resulting from waning development (*absteigende Entwicklung*), form when the uplift rate is less than the denudation rate. Later work has shown that valley-side shape depends not on the simple interplay of erosion rates and uplift rates, but on slope materials and the nature of slope-eroding processes.

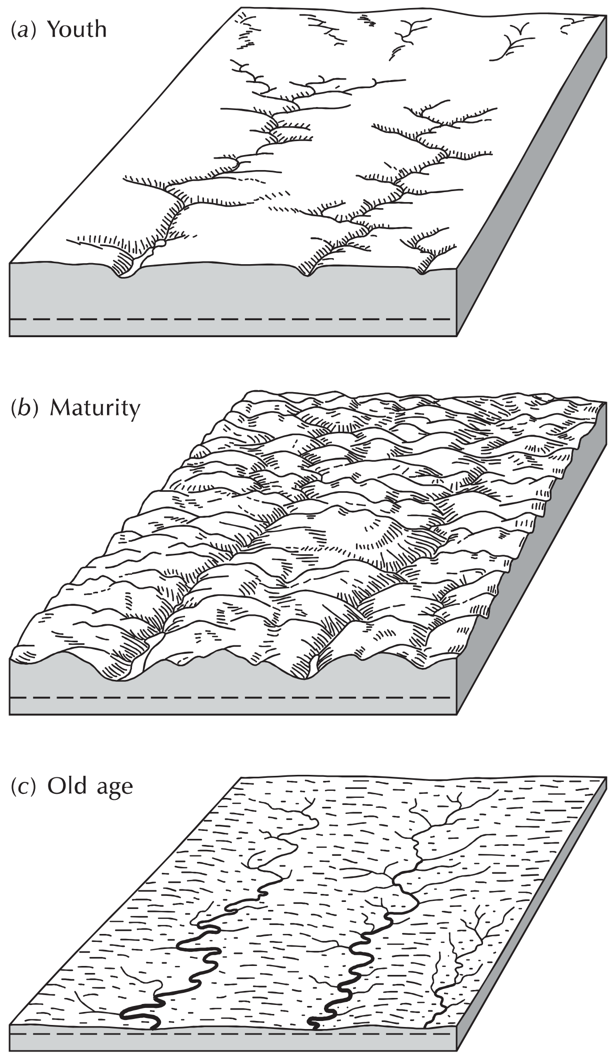


Figure 1.2 William Morris Davis's idealized 'geographical cycle' in which a landscape evolves through 'life-stages' to produce a peneplain. (a) Youth: a few 'consequent' streams (p. 135), V-shaped valley cross-sections, limited floodplain formation, large areas of poorly drained terrain between streams with lakes and marshes, waterfalls and rapids common where streams cross more resistant beds, stream divides broad and ill-defined, some meanders on the original surface. (b) Maturity: well-integrated drainage system, some streams exploiting lines of weak rocks, master streams have attained grade (p. 229), waterfalls, rapids, lakes, and marshes largely eliminated, floodplains common on valley floors and bearing meandering rivers, valley no wider than the width of meander belts, relief (difference in elevation between highest and lowest points) is at a maximum, hillslopes and valley sides dominate the landscape. (c) Old age: trunk streams more important again, very broad and gently sloping valleys, floodplains extensive and carrying rivers with broadly meandering courses, valleys much wider than the width of meander belts, areas between streams reduced in height and stream divides not so sharp as in the maturity stage, lakes, swamps, and marshes lie on the floodplains, mass-wasting dominates fluvial processes, stream adjustments to rock types now vague, extensive areas lie at or near the base level of erosion.
Source: Adapted from Holmes (1965, 473)

According to Penck's arguments, slopes may either recede at the original gradient or else flatten, according to circumstances. Many textbooks claim that Penck advocated 'parallel retreat of slopes', but this is a false belief (see Simons 1962). Penck (1953, 135–6) argued that a steep rock face would move upslope, maintaining its original gradient, but would soon be eliminated by a growing basal slope. If the cliff face was the scarp of a tableland, however, it would take a long time to disappear. He reasoned that a lower-angle slope, which starts growing from the bottom of the basal slope, replaces the basal slope. Continued slope replacement then leads to a flattening of slopes, with steeper sections formed during earlier stages of development sometimes surviving in summit areas (Penck 1953, 136–41). In short, Penck's complicated analysis predicted both **slope recession** and **slope decline**, a result that extends Davis's simple idea of **slope decline** (Figure 1.3). Field studies have confirmed that slope retreat is common in a wide range of situations. However, a slope that is actively eroded at its base (by a river or by the sea) may decline if the basal erosion should stop. Moreover, a tableland scarp retains its angle through parallel retreat until the erosion removes the protective cap rock, when slope decline sets in (Ollier and Tuddenham 1962).

Eduard Brückner and Albrecht Penck

Other early historical geomorphologists used geologically young sediments to interpret Pleistocene events. **Eduard Brückner** and **Albrecht Penck**'s (Walther's father) work on glacial effects on the Bavarian Alps and their forelands provided the first insights into the effects of the Pleistocene ice ages on relief (Penck and Brückner 1901–9). Their classic river-terrace sequence gave names to the main glacial stages – Donau, Gunz, Mindel, Riss, and Würm – and sired Quaternary geomorphology.

Modern historical geomorphology

Historical geomorphology has developed since Davis's time, and the interpretation of long-term changes of landscape no longer relies on the straitjacket of the geographical cycle. It relies now on various chronological analyses, particularly those based on stratigraphical studies of Quaternary sediments, and upon a much fuller appreciation of geomorphic and tectonic processes (e.g. Brown 1980). Observed stratigraphical relationships furnish relative chronologies, whilst absolute chronologies derive from sequences dated using historical records, radiocarbon analysis, dendrochronology, luminescence, palaeomagnetism, and so forth (p. 354). Such quantitative chronologies offer a means for calculating long-term rates of change in the landscape.

It is perhaps easiest to explain modern historical geomorphology by way of an example. Take the case of the river alluvium and colluvium that fills many valleys in countries bordering the Mediterranean Sea. Claudio Vita-Finzi (1969) pioneered research into the origin of the valley fills, concluding that almost all alluvium

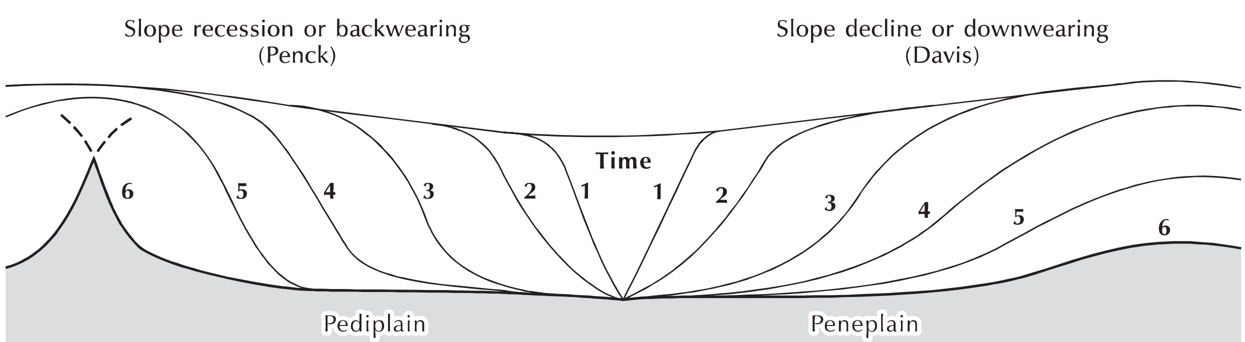


Figure 1.3 Slope recession, which produces a pediplain (p. 381) and slope decline, which produces a peneplain.
Source: Adapted from Gossman (1970)

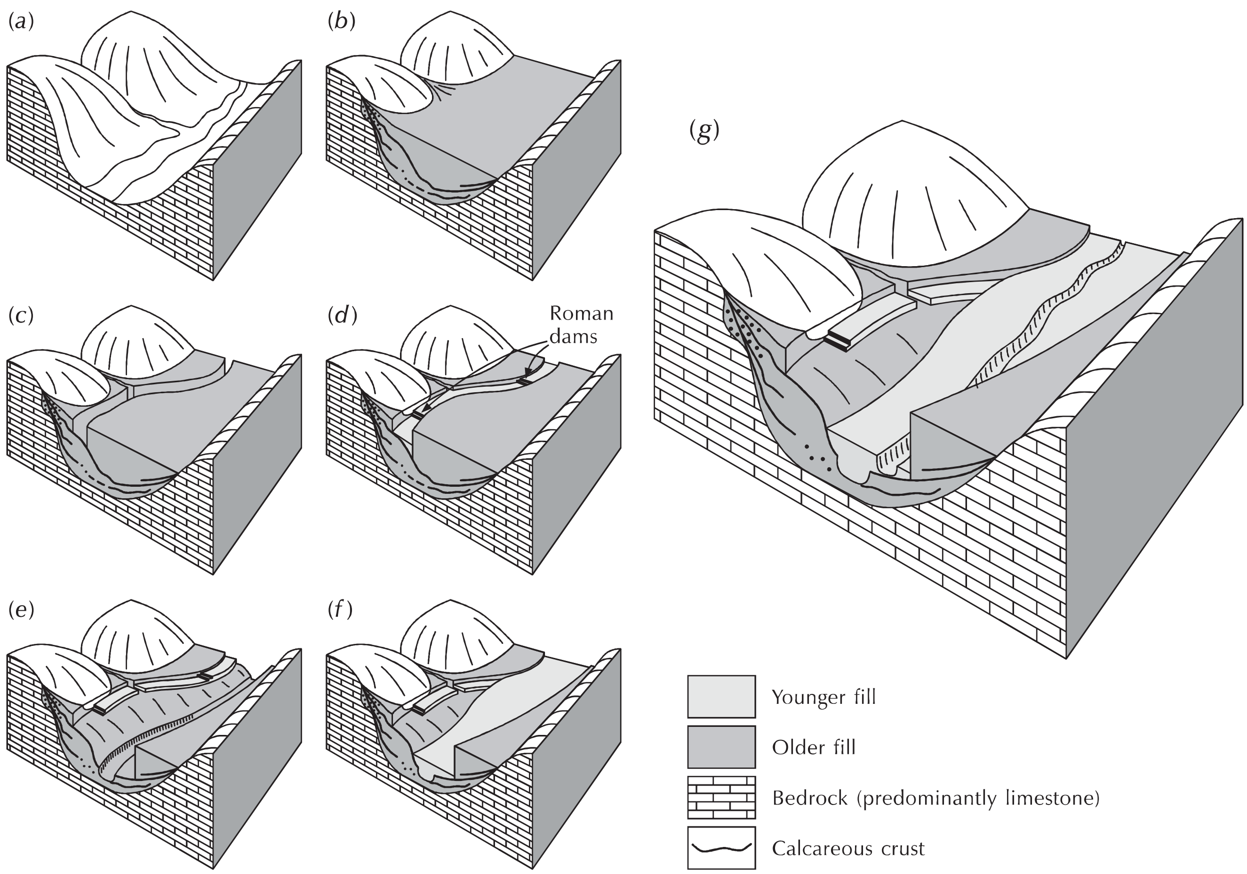


Figure 1.4 A reconstruction of the geomorphic history of a wadi in Tripolitania. (a) Original valley. (b) Deposition of Older Fill. (c) River cut into Older Fill. (d) Roman dams impound silt. (e) Rivers cut further into Older Fill and Roman alluvium. (f) Deposition of Younger Fill. (g) Present valley and its alluvial deposits.
Source: After Vita-Finzi (1969, 10)

and colluvium was laid down during two episodes of increased aggradation (times when deposition of sediment outstripped erosion). Figure 1.4 is a schematic reconstruction of the geomorphic history of a valley in Tripolitania (western Libya). The key to unlocking the history of the valleys in the area was datable archaeological material in the fluvial deposits. Vita-Finzi found three main deposits of differing ages. The oldest contains Palaeolithic implements and seems to have accumulated during the Pleistocene. Rivers cut into it between about 9,000 and 3,000 years ago. The second deposit accumulated behind dams built by Romans to store water and retain sediment. Late in the Empire, floodwaters breached or found a way around the dams and cut into the Roman alluvium. Rivers built up the third deposit, which contained Roman and earlier material as well as pottery and charcoal placing in the Medieval Period (AD 1200–1500), within the down-cut wadis. The deposition of this Younger Fill was followed by reduced alluviation and down-cutting through the fill.

Wider examination of alluvia in Mediterranean valleys allowed Vita-Finzi to recognize an Older Fill dating from the Pleistocene and a Younger Fill dating from about AD 500–1500. The Older Fill was deposited as a substantial body of colluvium (slope wash) under a 'periglacial' regime during the last glacial stage. The Younger Fill was a product of phases of erosion during the later Roman Imperial times, through the Dark Ages, and to the Middle Ages. Vita-Finzi believed it to be the result of increased erosion associated with the climate of the Medieval Warm Period or the Little Ice Age, a view supported by John Bintliff (1976, 2002). Other geomorphologists, including Karl Butzer (1980, 2005) and Tjierd van Andel and his co-workers (1986), favoured human activity as the chief cause, pointing to post-medieval deforestation and agricultural expansion into marginal environments. The matter is still open to debate (see p. 363).

Process geomorphology

Process geomorphology is the study of the processes responsible for landform development. In the modern era, the first process geomorphologist, carrying on the tradition started by Leonardo da Vinci (p. 5), was Grove Karl Gilbert. In his treatise on the Henry Mountains of Utah, USA, Gilbert discussed the mechanics of fluvial processes (Gilbert 1877), and later he investigated the transport of debris by running water (Gilbert 1914). Up to about 1950, when the subject grew apace, important contributors to process geomorphology included Ralph Alger Bagnold (p. 85), who considered the physics of blown sand and desert dunes, and Filip Hjulstrøm (p. 73), who investigated fluvial processes. After 1950, several 'big players' emerged that set process geomorphology moving apace. Arthur N. Strahler was instrumental in establishing process geomorphology, his 1952 paper called 'Dynamic basis of geomorphology' being a landmark publication. John T. Hack, developing Gilbert's ideas, prosecuted the notions of **dynamic equilibrium** and **steady state**, arguing that a landscape should attain a steady state, a condition in which land-surface form does not change despite material being added by tectonic uplift and removed by a constant set of geomorphic processes. And he contended that, in an erosional landscape, dynamic equilibrium prevails where all slopes, both hillslopes and river slopes, are adjusted to each other (cf. Gilbert 1877, 123–4; Hack 1960, 81), and 'the forms and processes are in a steady state of balance and may be considered as time independent' (Hack 1960, 85). Luna B. Leopold and M. Gordon Wolman made notable contributions to the field of fluvial geomorphology (e.g. Leopold *et al.* 1964). Stanley A. Schumm, another fluvial geomorphologist, refined notions of landscape stability to include **thresholds** and dynamically **metastable states** and made an important contribution to the understanding of timescales (p. 27). Stanley W. Trimble worked on historical and modern **sediment budgets** in small catchments (e.g. Trimble 1983). Richard J. Chorley brought process geomorphology to the UK and demonstrated the power of a **systems approach** to the subject.

Process geomorphologists have done their subject at least three great services. First, they have built up a database of process rates in various parts of the globe. Second, they have built increasingly refined models for predicting the short-term (and in some cases long-term) changes in landforms. Third, they have generated some enormously powerful ideas about stability and instability in geomorphic systems (see pp. 19–21).

Measuring geomorphic processes

Some geomorphic processes have a long record of measurement. The oldest year-by-year record is the flood levels of the River Nile in lower Egypt. Yearly readings at Cairo are available from the time of Muhammad, and some stone-inscribed records date from the first dynasty of the pharaohs, around 3100 BC. The amount of sediment annually carried down the Mississippi River was gauged during the 1840s, and the rates of modern denudation in some of the world's major rivers were estimated in the 1860s. The first efforts to measure weathering rates were made in the late nineteenth century. Measurements of the dissolved load of rivers enabled estimates of chemical denudation rates to be made in the first half of the twentieth century, and patchy efforts were made to widen the range of processes measured in the field. But it was the quantitative revolution in geomorphology, started in the 1940s, that was largely responsible for the measuring of process rates in different environments. Since about 1950, the attempts to quantify geomorphic processes in the field have grown fast. An early example is the work of Anders Rapp (1960), who tried to quantify all the processes active in a subarctic environment and assess their comparative significance. His studies enabled him to conclude that the most powerful agent of removal from the

Karkevagge drainage basin was running water bearing material in solution. An increasing number of hillslopes and drainage basins have been instrumented, that is, had measuring devices installed to record a range of geomorphic processes. The instruments used on hillslopes and in geomorphology generally are explained in several books (e.g. Goudie 1994). Interestingly, some of the instrumented catchments established in the 1960s have recently received unexpected attention from scientists studying global warming, because records lasting decades in climatically sensitive areas – high latitudes and high altitudes – are invaluable. However, after half a century of intensive field measurements, some areas, including Europe and North America, still have better coverage than other areas. And field measurement programmes should ideally be ongoing and work on as fine a resolution as practicable, because rates measured at a particular place may vary through time and may not be representative of nearby places.

Modelling geomorphic processes

Since the 1960s and 1970s, process studies have been largely directed towards the construction of models for predicting short-term changes in landforms, that is, changes happening over human timescales. Such models have drawn heavily on soil engineering, for example in the case of slope stability, and hydraulic engineering in the cases of flow and sediment entrainment and deposition in rivers. Nonetheless, some geomorphologists, including Michael J. Kirkby and Jonathan D. Phillips, have carved out a niche for themselves in the modelling department. An example of a geomorphic model is shown in Figure 1.5 (see also p. 22).

Process studies and global environmental change

With the current craze for taking a global view, process geomorphology has found natural links with other Earth and life sciences. Main thrusts of research investigate (1) energy and mass fluxes and (2) the response of landforms to climate, hydrology, tectonics, and land use (Slaymaker 2000b, 5). The focus on mass and energy fluxes explores the short-term links between land-surface systems and climate that are forged through the storages and movements of energy, water, biogeochemicals, and sediments. Longer-term and broader-scale interconnections between landforms and climate, water budgets, vegetation cover, tectonics, and human activity are a focus for process geomorphologists who take a historical perspective and investigate the causes and effects of changing processes regimes during the Quaternary.

Applied geomorphology

Applied geomorphology studies the interactions of humans with landscapes and landforms. Process geomorphologists, armed with their models, have contributed to the investigation of worrying problems associated with the human impacts on landscapes. They have studied coastal erosion and beach management (e.g. Bird 1996; Viles and Spencer 1996), soil erosion, the weathering of buildings, landslide protection, river management and river channel restoration (e.g. Brookes and Shields 1996), and the planning and design of landfill sites (e.g. Gray 1993). Other process geomorphologists have tackled general applied issues. *Geomorphology in Environmental Planning* (Hooke 1988), for example, considered the interaction between geomorphology and public policies, with contributions on rural land-use and soil erosion, urban land-use, slope management, river management, coastal management, and policy formulation. *Geomorphology in Environmental Management* (Cooke 1990), as its title suggests, looked at the role played by geomorphology in management aspects of the environment. *Geomorphology and Land Management in a Changing Environment* (McGregor and Thompson 1995) focused upon problems of managing land against a background of environmental change. The conservation of ancient and modern landforms is an expanding aspect of applied geomorphology.

Three aspects of applied geomorphology have been brought into a sharp focus by the impending environmental change associated with global warming (Slaymaker 2000b) and illustrate the value of geomorphological know-how. First, applied geomorphologists are ideally placed to work on the mitigation of natural hazards of geomorphic origin, which may well increase in magnitude and frequency during the twenty-first century

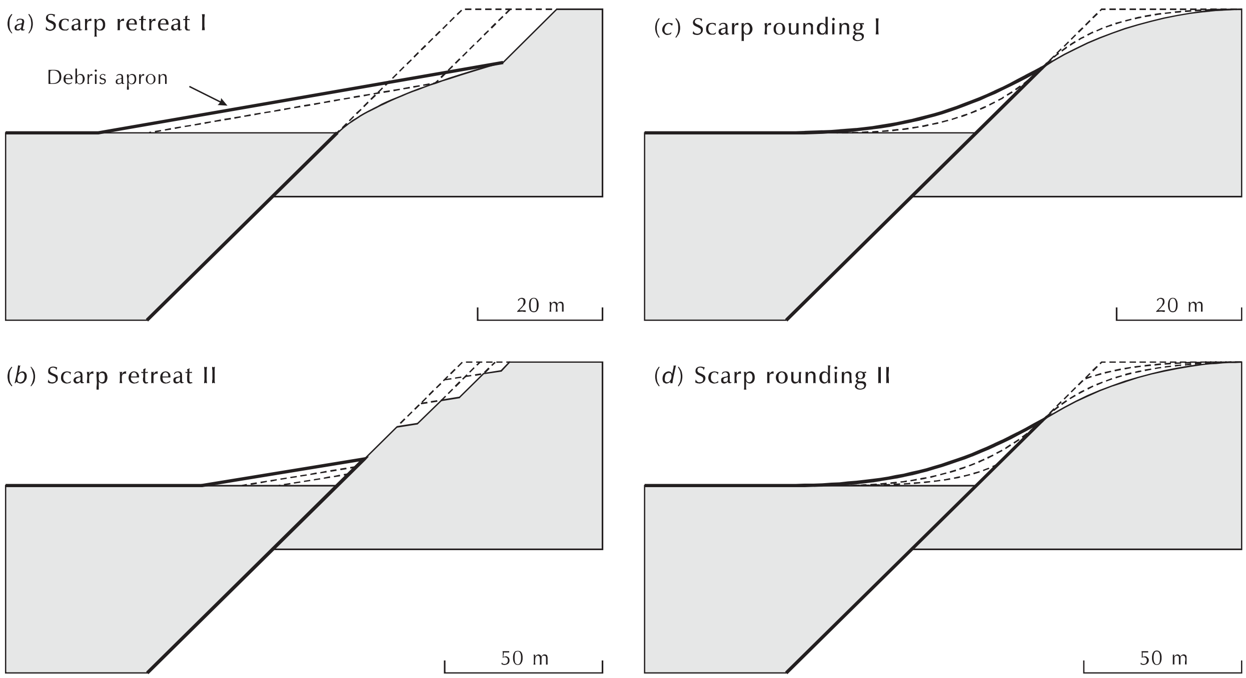


Figure 1.5 Example of a geomorphic model: the predicted evolution of a fault scarp according to assumptions made about slope processes. (a) Parallel scarp retreat with deposition of debris at the base. The scarp is produced by a single movement along the fault. (b) Parallel scarp retreat with deposition at the base. The scarp is produced by four separate episodes of movement along the fault. In cases (a) and (b) it is assumed that debris starts to move downslope once a threshold angle is reached and then comes to rest where the scarp slope is less than the threshold angle. Allowance is made for the packing density of the debris and for material transported beyond the debris apron. (c) Rounding of a fault scarp that has been produced by one episode of displacement along the fault. (d) Rounding of a fault scarp that has been produced by four separate episodes of movement along the fault. In cases (c) and (d), it is assumed that the volume of debris transported downslope is proportional to the local slope gradient.
Source: Adapted from Nash (1981)

and beyond. Landslides and debris flows may become more common, soil erosion may become more severe and the sediment load of some rivers increase, some beaches and cliffs may erode faster, coastal lowlands may become submerged, and frozen ground in the tundra environments may thaw. Applied geomorphologists can address all these potentially damaging changes. Second, a worrying aspect of global warming is its effect on natural resources – water, vegetation, crops, and so on. Applied geomorphologists, equipped with such techniques as terrain mapping, remote sensing, and geographical information systems, can contribute to environmental management programmes. Third, applied geomorphologists are able to translate the predictions of global and regional temperature rises into predictions of critical boundary changes, such as the poleward shift of the permafrost line and the tree-line, which can then guide decisions about tailoring economic activity to minimize the effects of global environmental change.

Other geomorphologies

There are many other kinds of geomorphology, including tectonic geomorphology, submarine geomorphology, climatic geomorphology, and planetary geomorphology.

Tectonic geomorphology is the study of the interplay between tectonic and geomorphic processes in regions where the Earth's crust actively deforms. Advances in the measurement of rates and in the understanding of the physical basis of tectonic and geomorphic processes have revitalized it as a field of enquiry. It is a stimulating and highly integrative field that uses techniques and data drawn from studies of geomorphology, seismology, geochronology, structure, geodesy, and Quaternary climate change (e.g. Burbank and Anderson 2001).

Submarine geomorphology deals with the form, origin, and development of features of the sea floor. Submarine landforms cover about 71 per cent of the Earth's surface, but are mostly less well studied than their terrestrial counterparts. In shallow marine environments, landforms include ripples, dunes, sand waves, sand ridges, shorelines, and subsurface channels. In the continental slope transition zone are submarine canyons and gullies, inter-canyon areas, intraslope basins, and slump and slide scars. The deep marine environment contains varied landforms, including trench and basin plains, trench fans, sediment wedges, abyssal plains, distributary channels, and submarine canyons.

Planetary geomorphology is the study of landforms on planets and large moons with a solid crust, for example Venus, Mars, and some moons of Jupiter and Saturn. It is a thriving branch of geomorphology (e.g. Howard 1978; Baker 1981; Grant 2000; Irwin *et al.* 2005). Surface processes on other planets and their satellites depend materially on their mean distance from the Sun, which dictates the annual receipt of solar energy, on their rotational period, and on the nature of the planetary atmosphere. Observed processes include weathering, aeolian activity, fluvial activity, glacial activity, and mass wasting.

Climatic geomorphology rests on the not universally accepted observation that each climatic zone (tropical, arid, temperate for example) engenders a distinctive suite of landforms (e.g. Tricart and Cailleux 1972; Büdel 1982). Climate does strongly influence geomorphic processes, but it is doubtful that the set of geomorphic processes within each climatic zone creates characteristic landforms. The current consensus is that, owing to climatic and tectonic change, the climatic factor in landform development is more complicated than climatic geomorphologists have on occasions suggested (cf. p. 389–90).

FORM

The two main approaches to form in geomorphology are description (field description and morphological mapping) and mathematical representation (geomorphometry).

Field description and morphological mapping

The only way fully to appreciate landforms is to go into the field and see them. Much can be learnt from the now seemingly old-fashioned techniques of field description, field sketching, and map reading and map making.

The mapping of landforms is an art (see Dackombe and Gardiner 1983, 13–20, 28–41; Evans 1994). Landforms vary enormously in shape and size. Some, such as karst depressions and volcanoes, may be represented as points. Others, such as faults and rivers, are linear features that are best depicted as lines. In other cases, areal properties may be of prime concern and suitable means of spatial representation must be employed. Morphological maps capture areal properties. **Morphological mapping** attempts to identify basic landform units in the field, on aerial photographs, or on maps. It sees the ground surface as an assemblage of landform elements. **Landform elements** are recognized as simply curved geometric surfaces lacking inflections (complicated kinks) and are considered in relation to upslope, downslope, and lateral elements. They go by a plethora of names – facets, sites, land elements, terrain components, and facies. The 'site' (Linton 1951) was an elaboration of the 'facet' (Wooldridge 1932), and involved altitude, extent, slope, curvature, ruggedness, and relation to the water table. The other terms were coined in the 1960s (see Speight 1974). Figure 1.6 shows the land surface of Longdendale in the Pennines, England, represented as a morphological map. The map combines landform elements derived from a nine-unit land-surface model (p. 169) with depictions of deep-seated mass movements and superficial mass movements. Digital elevation models lie within the ambits of landform morphometry and are dealt with below. They have greatly extended, but by no means replaced, the classic

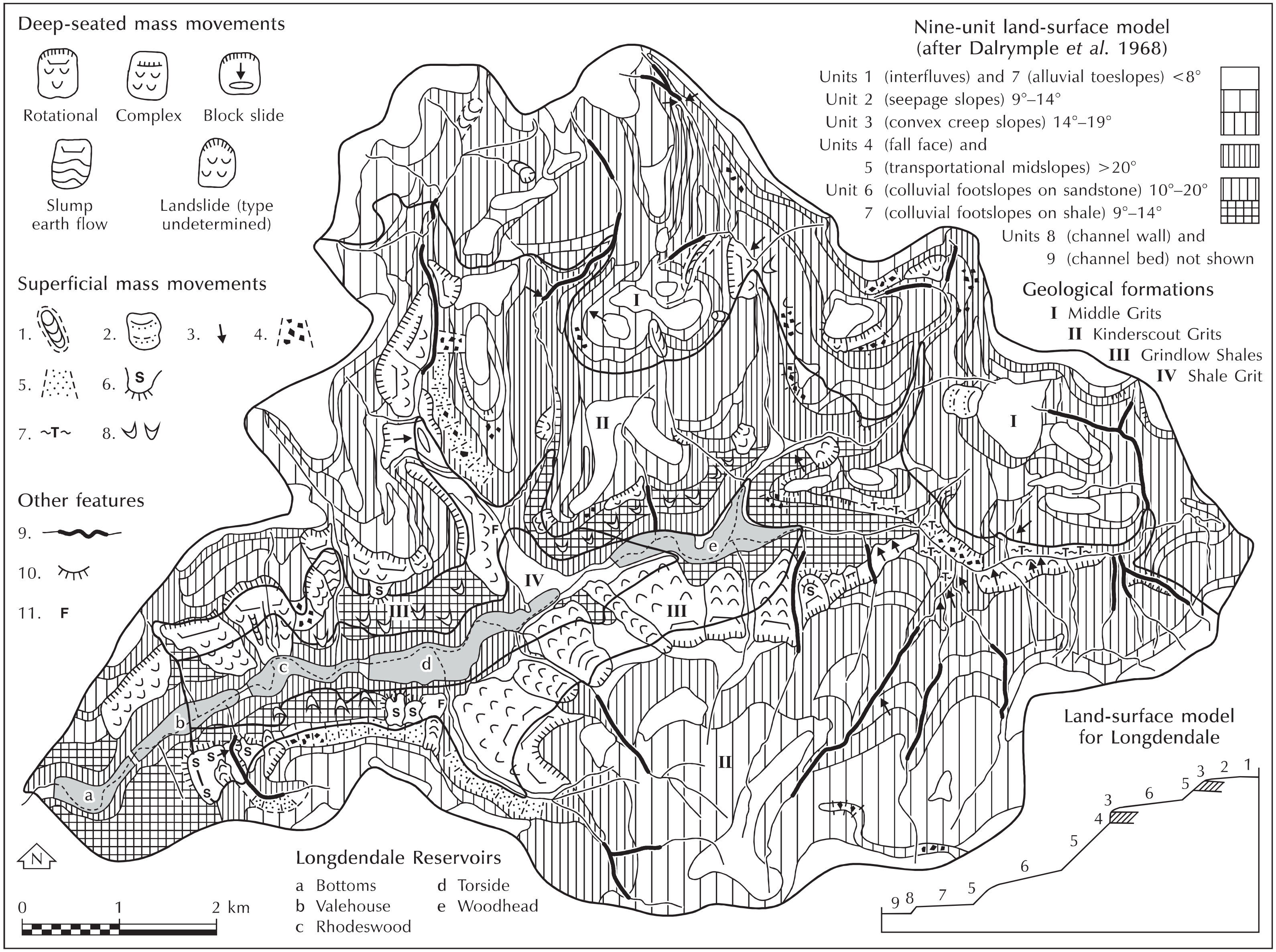


Figure 1.6 Morphological map of Longdendale, north Derbyshire, England. The map portrays units of a nine-unit land-surface model, types of mass movement, and geological formations. The superficial mass movements are: 1 mudflow, earthflow, or peat burst; 2 Soil slump; 3 Minor soil slump; 4 Rockfall; 5 Scree; 6 Solifluction lobe; 7 Terracettes; 8 Soil creep or block creep and soliflucted material. The other features are: 9 Incised stream; 10 Rock cliff; 11 Valley-floor alluvial fan.
Source: After Johnson (1980)

work on landform elements and their descriptors as prosecuted by the morphological mappers.

Geomorphometry

A branch of geomorphology – **landform morphometry** or **geomorphometry** – studies quantitatively the form of the land surface. Geomorphometry in the modern era is traceable to the work of Alexander von Humboldt and Carl Ritter in the early and mid-nineteenth century (see Pike 1999). It had a strong post-war tradition in North America and the UK, and it has been 'reinvented' with the advent of remotely sensed images and Geographical Information Systems (GIS) software. The contributions of geomorphometry to geomorphology and cognate fields are legion. Geomorphometry is an important component of terrain analysis and surface modelling. Its specific applications include measuring the morphometry of continental ice surfaces, characterizing glacial troughs, mapping sea-floor terrain types, guiding missiles, assessing soil erosion, analysing wildfire propagation, and mapping ecoregions (Pike 1995, 1999). It also contributes to engineering, transportation, public works, and military operations.

Digital elevation models

The resurgence of geomorphometry since the 1970s is in large measure due to two developments. First is the light-speed development and use of **GIS**, which allow input, storage, and manipulation of digital data representing spatial and aspatial features of the Earth's surface. Second is the development of **Electronic Distance Measurement** (**EDM**) in surveying and, more recently, the **Global Positioning System** (**GPS**), which made the very time-consuming process of making large-scale maps much quicker and more fun. The spatial form of surface topography is modelled in several ways. Digital representations are referred to as either **Digital Elevation Models** (**DEMs**) or **Digital Terrain Models** (**DTMs**). A DEM is 'an ordered array of numbers that represent the spatial distribution of elevations above some arbitrary datum in a landscape' (Moore *et al.* 1991, 4). DTMs are 'ordered arrays of numbers that represent the spatial distribution of terrain attributes' (Moore *et al.* 1991, 4). DEMs are, therefore, a subset of DTMs. Topographic elements of a landscape can be computed directly from a DEM (p. 170). Further details of DEMs and their applications are given in several recent books (e.g. Wilson and Gallant 2000; Huggett and Cheesman 2002).

PROCESS

Geomorphic systems

Process geomorphologists commonly adopt a **systems approach** to their subject. To illustrate what this approach entails, take the example of a hillslope system. A hillslope extends from an interfluve crest, along a valley side, to a sloping valley floor. It is a system insofar as it consists of things (rock waste, organic matter, and so forth) arranged in a particular way. The arrangement is seemingly meaningful, rather than haphazard, because it is explicable in terms of physical processes (Figure 1.7). The 'things' of which a hillslope is composed may be described by such variables as particle size, soil moisture content, vegetation cover, and slope angle. These variables, and many others, interact to form a regular and connected whole: a hillslope, and the mantle of debris on it, records a propensity towards reciprocal adjustment among a complex set of variables. The complex set of variables include rock type, which influences weathering rates, the geotechnical properties of the soil, and rates of infiltration; climate, which influences slope hydrology and so the routing of water over and through the hillslope mantle; tectonic activity, which may alter baselevel; and the geometry of the hillslope, which, acting mainly through slope angle and distance from the divide, influences the rates of processes such as landsliding, creep, solifluction, and wash. Change in any of the variables will tend to cause a readjustment of hillslope form and process.

Isolated, open, and closed systems

Systems of all kinds are open, closed, or isolated according to how they interact, or do not interact, with their surroundings (Huggett 1985, 5–7). Traditionally, an **isolated system** is a system that is completely cut off from its surroundings and that cannot therefore import or

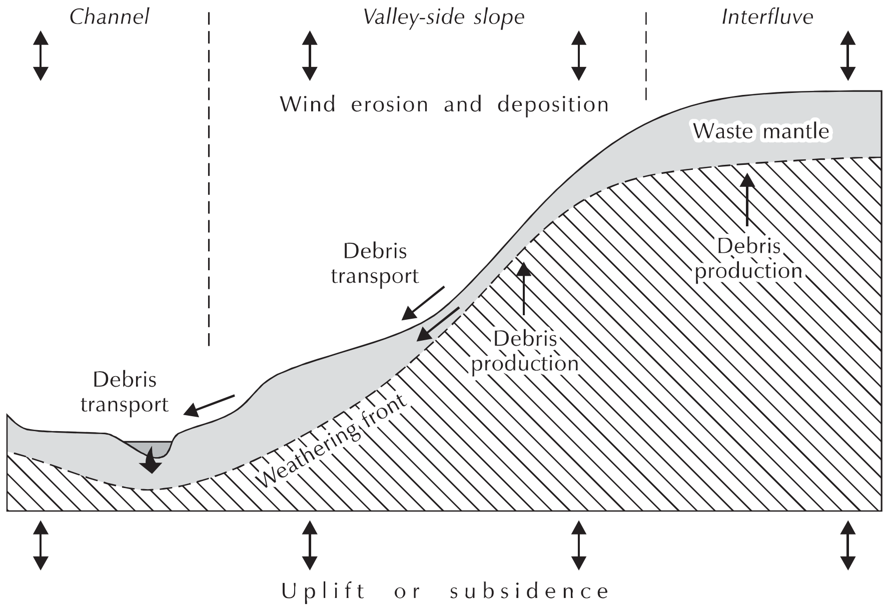


Figure 1.7 A hillslope as a system, showing storages (waste mantle), inputs (e.g. wind deposition and debris production), outputs (e.g. wind erosion), throughputs (debris transport), and units (channel, valley-side slope, interfluve).

export matter or energy. A **closed system** has boundaries open to the passage of energy but not of matter. An **open system** has boundaries across which energy and materials may move. All geomorphic systems, including hillslopes, may be thought of as open systems as they exchange energy and matter with their surroundings.

Internal and external system variables

Any geomorphic system has **internal** and **external variables**. Take a drainage basin. Soil wetness, streamflow, and other variables lying inside the system are endogenous or internal variables. Precipitation, solar radiation, tectonic uplift, and other such variables originating outside the system and affecting drainage basin dynamics are exogenous or external variables. Interestingly, all geomorphic systems can be thought of as resulting from a basic antagonism between **endogenic** (**tectonic** and **volcanic**) processes driven by geological forces and **exogenic** (**geomorphic**) processes driven by climatic forces (Scheidegger 1979). In short, tectonic processes create land, and climatically influenced weathering and erosion destroy it. The events between the creation and the final destruction are what fascinate geomorphologists.

Systems are mental constructs and have been defined in various ways. Two conceptions of systems are important in geomorphology: systems as process and form structures, and systems as simple and complex structures (Huggett 1985, 4–5, 17–44).

Geomorphic systems as form and process structures

Three kinds of geomorphic system may be identified: form systems, process systems, and form and process systems.

1 Form systems. **Form** or **morphological systems** are defined as sets of form variables that are deemed to interrelate in a meaningful way in terms of system origin or system function. Several measurements could be made to describe the form of a hillslope system. Form elements would include measures of anything on a hillslope that has size, shape, or physical properties. A simple characterization of hillslope form is

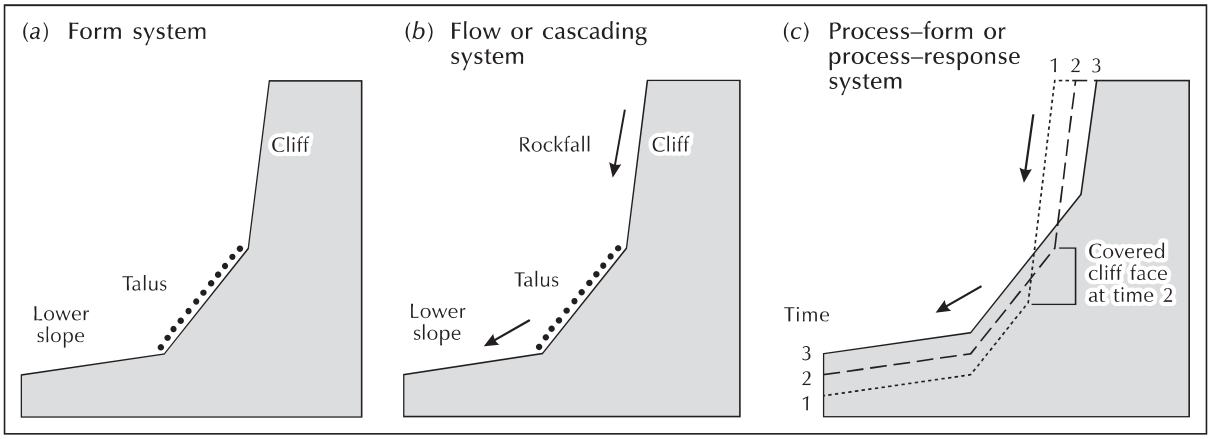


Figure 1.8 A cliff and talus slope viewed as (a) a form system, (b) a flow or cascading system, and (c) a process–form or process–response system. Details are given in the text.

shown in Figure 1.8a, which depicts a cliff with a talus slope at its base. All that could be learnt from this 'form system' is that the talus lies below the cliff; no causal connections between the processes linking the cliff and talus slope are inferred. Sophisticated characterizations of hillslope and land-surface forms may be made using digital terrain models.

2 Process systems. **Process systems**, which are also called **cascading** or **flow systems**, are defined as 'interconnected pathways of transport of energy or matter or both, together with such storages of energy and matter as may be required' (Strahler 1980, 10). An example is a hillslope represented as a store of materials: weathering of bedrock and wind deposition add materials to the store, and erosion by wind and fluvial erosion at the slope base removes materials from the store. The materials pass through the system and in doing so link the morphological components. In the case of the cliff and talus slope, it could be assumed that rocks and debris fall from the cliff and deliver energy and rock debris to the talus below (Figure 1.8b).

3 Form and process systems. **Process–form systems**, also styled **process–response systems**, are defined as an energy-flow system linked to a form system in such a way that system processes may alter the system form and, in turn, the changed system form alters the system processes. A hillslope may be viewed in this way with slope form variables and slope process variables interacting. In the cliff-and-talus example, rock falling off the cliff builds up the talus store (Figure 1.8c). However, as the talus store increases in size, so it begins to bury the cliff face, reducing the area that supplies debris. In consequence, the rate of talus growth diminishes and the system changes at an ever-decreasing rate. The process described is an example of negative feedback, which is an important facet of many process–form systems (Box 1.2).

Geomorphic systems as simple or complex structures

Three main types of system are recognized under this heading: simple systems, complex but disorganized systems, and complex and organized systems.

1 **Simple systems**. The first two of these types have a long and illustrious history of study. Since at least the seventeenth-century revolution in science, astronomers have referred to a set of heavenly bodies connected together and acting upon each other according to certain laws as a system. The Solar System is the Sun and its planets. The Uranian

Box 1.2

NEGATIVE AND POSITIVE FEEDBACK

Negative feedback is said to occur when a change in a system sets in motion a sequence of changes that eventually neutralize the effects of the original change, so stabilizing the system. An example occurs in a drainage basin system, where increased channel erosion leads to a steepening of valley-side slopes, which accelerates slope erosion, which increases stream bed-load, which reduces channel erosion (Figure 1.9a). The reduced channel erosion then stimulates a sequence of events that stabilizes the system and counteracts the effects of the original change. Some geomorphic systems also display **positive feedback** relationships characterized by an original change being magnified and the system being made unstable. An example is an eroding hillslope where the slope erosion causes a reduction in infiltration capacity of water, which increases the amount of surface runoff, which promotes even more slope erosion (Figure 1.9b). In short, a 'vicious circle' is created, and the system, being unstabilized, continues changing.

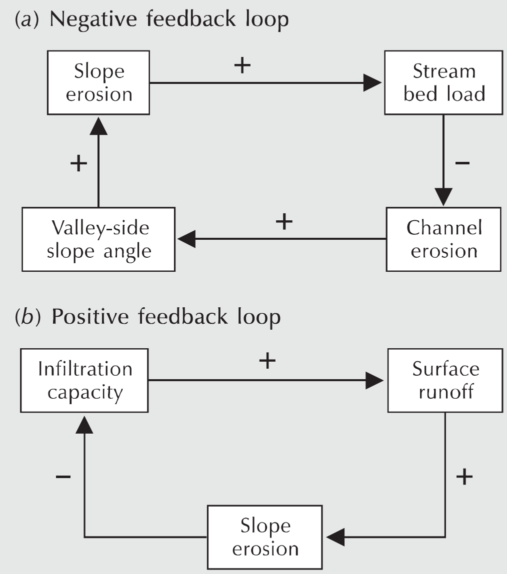


Figure 1.9 Feedback relationships in geomorphic systems. (a) Negative feedback in a valley-side slope–stream system. (b) Positive feedback in an eroding hillslope system. Details of the relationships are given in the text.

system is Uranus and its moons. These structures may be thought of as simple systems. In geomorphology, a few boulders resting on a talus slope may be thought of as a simple system. The conditions needed to dislodge the boulders, and their fate after dislodgement, can be predicted from mechanical laws involving forces, resistances, and equations of motion, in much the same way that the motion of the planets around the Sun can be predicted from Newtonian laws.

2 In a **complex but disorganized system**, a vast number of objects are seen to interact in a weak and haphazard way. An example is a gas in a jar. This system might comprise upward of 10^{23} molecules colliding with each other. In the same way, the countless individual particles in a hillslope mantle could be regarded as a complex but rather disorganized system. In both the gas and the hillslope mantle, the interactions are somewhat haphazard and far too numerous to study individually, so aggregate measures must be employed (see Huggett 1985, 74–7; Scheidegger 1991, 251–8).

3 In a third and later conception of systems, objects are seen to interact strongly with one another to form **systems of a complex and organized nature**. Most biological and ecological systems are of this kind. Many structures in geomorphology display high degrees of regularity and rich connections, and may be thought of as complexly organized systems. A hillslope represented as a process–form system could be placed into this category. Other examples include soils, rivers, and beaches.

Geomorphic system dynamics: equilibrium and steady state

As defined by John T. Hack, a steady-state landscape is one in which land-surface form stays the same despite tectonic uplift adding material and a constant set of geomorphic processes removing it. An erosional landscape in dynamic equilibrium arises when all slopes, both hillslopes and river slopes, are adjusted to each other (p. 10). In practice, this early notion of dynamic equilibrium was open to question (e.g. Ollier 1968) and difficult to apply to landscapes. In consequence, other forms of equilibrium were advanced (Howard 1988) (Figure 1.10). Of these, **dynamic metastable equilibrium** has proved to be salutary. It suggests that, once perturbed by

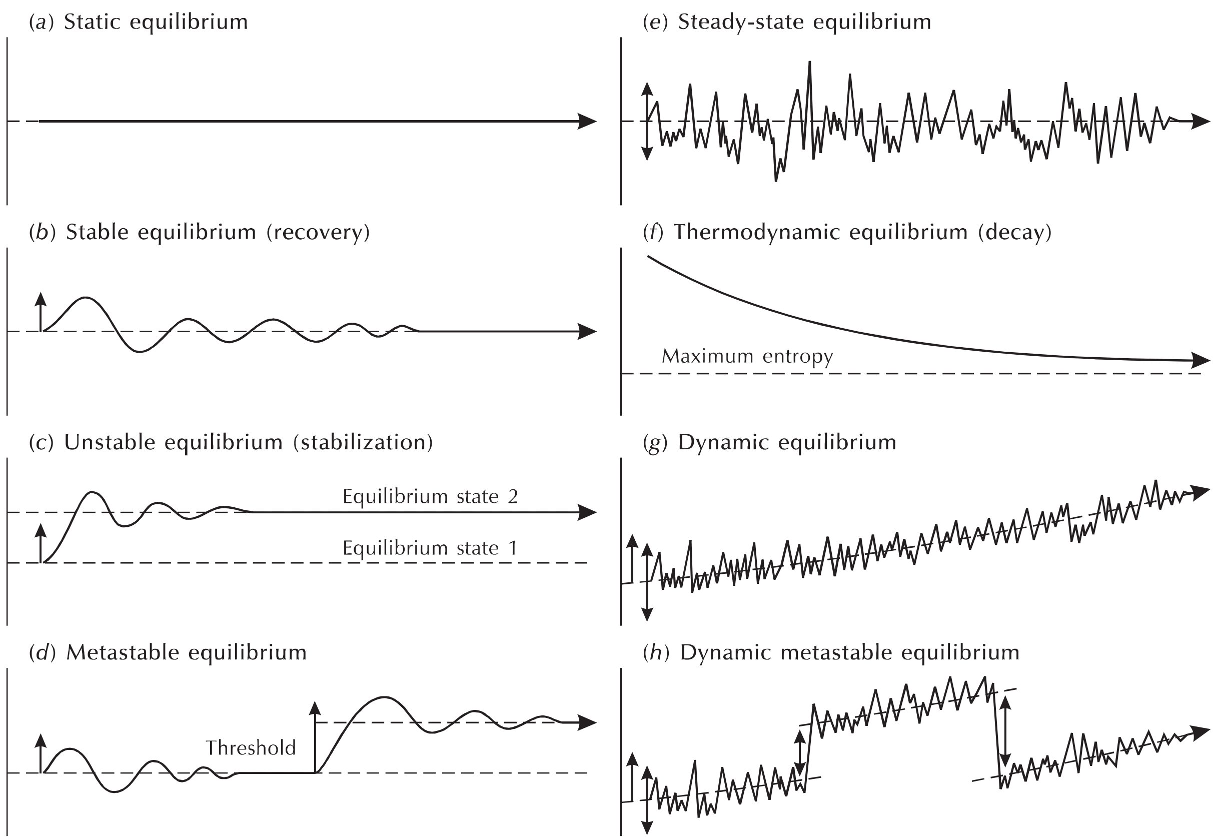


Figure 1.10 Types of equilibrium in geomorphology. (a) Static equilibrium occurs when a system is in balance over a time period and no change in state occurs. (b) Stable equilibrium records a tendency to revert to a previous state after a small disturbance. (c) Unstable equilibrium occurs when a small disturbance forces a system towards a new equilibrium state where stabilization occurs. (d) Metastable equilibrium arises when a system crosses an internal or external system threshold (p. 20), so driving it to a new state. (e) Steady state equilibrium obtains when a system constantly fluctuates about a mean equilibrium state. (f) Thermodynamic equilibrium is the tendency of some systems towards a state of maximum entropy, as in the gradual dissipation of heat by the Universe and its possible eventual 'heat death' and in the reduction of a mountain mass to a peneplain during a prolonged period of no uplift. (g) Dynamic equilibrium may be thought of as balanced fluctuations about a mean state that changes in a definite direction (a trending mean).
(h) Dynamic metastable equilibrium combines dynamic and metastable tendencies, with balanced fluctuations about a trending mean flipping to new trending mean values when thresholds are crossed.
Source: After Chorley and Kennedy (1971, 202)

Box 1.3

THRESHOLDS

A threshold separates different states of a system. It marks some kind of transition in the behaviour, operation, or state of a system. Everyday examples abound. Water in a boiling kettle crosses a temperature threshold in changing from a liquid to a gas. Similarly, ice taken out of a refrigerator and placed upon a table in a room with an air temperature of 10°C will melt because a temperature threshold has been crossed. In both examples, the huge differences in state – liquid water to water vapour, and solid water to liquid water – may result from tiny changes of temperature. Many geomorphic processes operate only after the crossing of a threshold. Landslides, for instance, require a critical slope angle, all other factors being constant, before they occur. Stanley A. Schumm (1979) made a powerful distinction between **external** and **internal system thresholds**. A geomorphic system will not cross an external threshold unless it is forced to do so by a change in an external variable. A prime example is the response of a geomorphic system to climatic change. Climate is the external variable. If, say, runoff were to increase beyond a critical level, then the geomorphic system might suddenly respond by reorganizing itself into a new state. No change in an external variable is required for a geomorphic system to cross an internal threshold. Rather, some chance fluctuation in an internal variable within a geomorphic system may take a system across an internal threshold and lead to its reorganization. This appears to happen in some river channels where an initial disturbance by, say, overgrazing in the river catchment triggers a complex response in the river channel: a complicated pattern of erosion and deposition occurs with phases of alluviation and downcutting taking place concurrently in different parts of the channel system (see below).

environmental changes or random internal fluctuations that cause the crossing of internal **thresholds** (Box 1.3), a landscape will respond in a complex manner (Schumm 1979). A stream, for instance, if it should be forced away from a steady state, will adjust to the change. However, the nature of the adjustment may vary in different parts of the stream and at different times. Douglas Creek in western Colorado, USA, was subject to overgrazing during the 'cowboy era' (Womack and Schumm 1977). It has been cutting into its channel bed since about 1882. The manner of incision has been complex, with discontinuous episodes of downcutting interrupted by phases of deposition, and with the erosion–deposition sequence varying from one cross-section to another. Trees have been used to date terraces at several locations. The terraces are unpaired (p. 236), which is not what would be expected from a classic case of river incision, and they are discontinuous in a downstream direction. This kind of study serves to dispel for ever the simplistic cause-and-effect view of landscape evolution in which change is seen as a simple response to an altered input. It shows that landscape dynamics may involve abrupt and discontinuous behaviour involving flips between quasi-stable states as system thresholds are crossed.

The latest views on landscape stability (or lack of it) come from the field of **dynamic systems theory**, which embraces the buzzwords **complexity** and **chaos**. The argument runs that steady states in the landscape may be rare because landscapes are inherently unstable. This is because any process that reinforces itself keeps the system changing through a positive feedback circuit and readily disrupts any balance obtaining in a steady state. This idea is formalized as an 'instability principle', which recognizes that, in many landscapes, accidental deviations from a 'balanced' condition tend to be self-reinforcing (Scheidegger 1983). This explains why cirques tend to grow, sinkholes increase in size, and longitudinal mountain valley profiles become stepped. The intrinsic instability of landscapes is borne out by mathematical analyses that point to the chaotic nature of much landscape change (e.g. Phillips 1999; Scheidegger 1994). Jonathan D. Phillips's (1999, 139–46) investigation into the nature of Earth surface systems, which includes geomorphic systems, is particularly revealing and will be discussed in the final chapter.

Magnitude and frequency

Interesting debates centre on the variations in process rates through time. The 'tame' end of this debate concerns arguments over **magnitude** and **frequency** (Box 1.4), the pertinent question here being which events perform the most geomorphic work: small and infrequent events, medium and moderately frequent events, or big but rare events? The first work on this issue concluded, albeit provisionally until further field work was carried out, that events occurring once or twice a year perform most geomorphic work (Wolman and Miller 1960). Some later work has highlighted the geomorphic significance of rare events. Large-scale anomalies in atmospheric circulation systems very occasionally produce short-lived superfloods that have long-term effects on landscapes (Baker 1977, 1983; Partridge and Baker 1987). Another study revealed that low-frequency, high-magnitude events greatly affect stream channels (Gupta 1983).

The 'wilder' end engages hot arguments over **gradualism** and **catastrophism** (Huggett 1989, 1997a, 2006). The crux of the gradualist–catastrophist debate is the seemingly innocuous question: have the present rates of geomorphic processes remained much the same throughout Earth surface history? Gradualists claim that process

Box 1.4

MAGNITUDE AND FREQUENCY

As a rule of thumb, bigger floods, stronger winds, higher waves, and so forth occur less often than their smaller, weaker, and lower counterparts. Indeed, graphs showing the relationship between the frequency and magnitude of many geomorphic processes are right-skewed, which means that a lot of low-magnitude events occur in comparison with the smaller number of high-magnitude events, and a very few very high-magnitude events. The frequency with which an event of a specific magnitude occurs is expressed as the **return period** or **recurrence interval**. The recurrence interval is calculated as the average length of time between events of a given magnitude. Take the case of river floods. Observations may produce a dataset comprising the maximum discharge for each year over a period of years. To compute the **flood–frequency relationships**, the peak discharges are listed according to magnitude, with the highest discharge first. The recurrence interval is then calculated using the equation

$$T = \frac{n+1}{m}$$

where T is the recurrence interval, n is the number of years of record, and m is the magnitude of the flood (with $m = 1$ at the highest recorded discharge). Each flood is then plotted against its recurrence interval on Gumbel graph paper and the points connected to form a frequency curve. If a flood of a particular magnitude has a recurrence interval of 10 years, it would mean that there is a 1-in-10 (10 per cent) chance that a flood of this magnitude (2,435 cumecs in the Wabash River example shown in Figure 1.11) will occur in any year. It also means that, on average, one such flood will occur every 10 years. The magnitudes of 5-year, 10-year, 25-year, and 50-year floods are helpful for engineering work, flood control, and flood alleviation. The 2.33-year flood ($Q_{2.33}$) is the mean annual flood (1,473 cumecs in the example), the 2.0-year flood ($Q_{2.0}$) is the median annual flood (not shown), and the 1.58-year flood ($Q_{1.58}$) is the most probable flood (1,133 cumecs in the example).

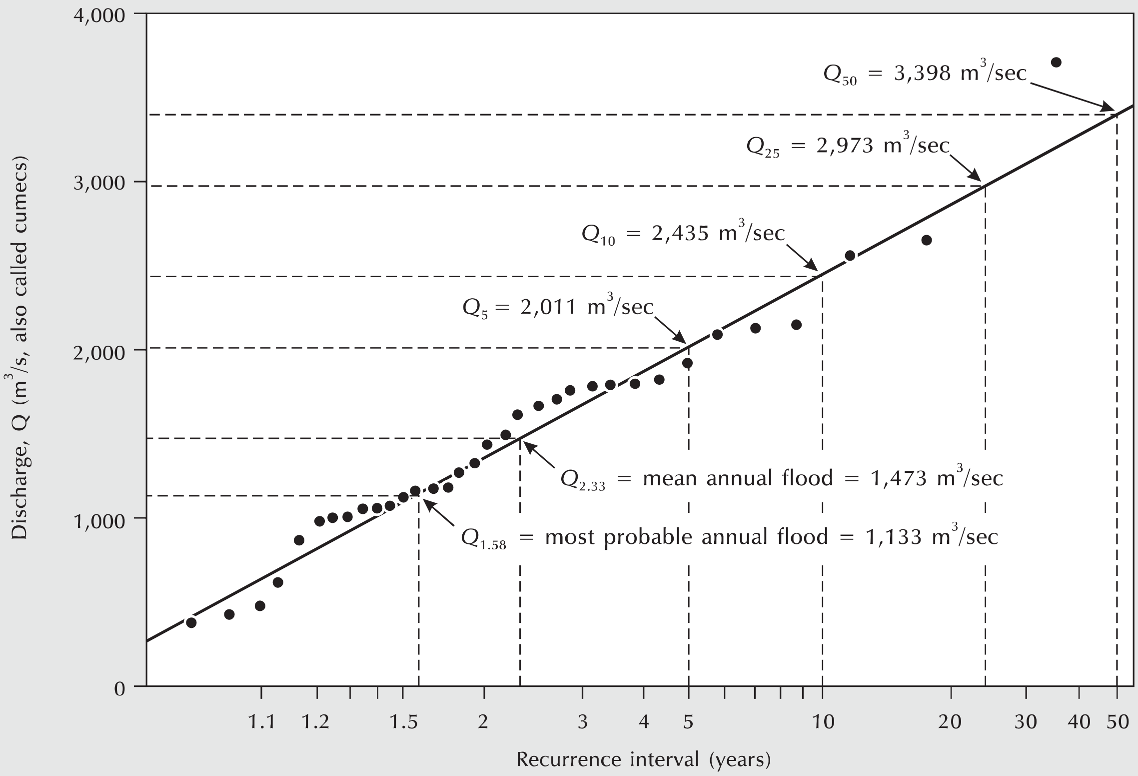


Figure 1.11 Magnitude–frequency plot of annual floods on the Wabash River, at Lafayette, Indiana, USA. See text for details.
Source: Adapted from Dury (1969)

rates have been uniform in the past, not varying much beyond their present levels. Catastrophists make the counterclaim that the rates of geomorphic processes have differed in the past, and on occasions some of them have acted with suddenness and extreme violence, pointing to the effects of massive volcanic explosions, the impacts of asteroids and comets, and the landsliding of whole mountainsides into the sea. The dichotomy between gradualists and catastrophists polarizes the spectrum of possible rates of change. It suggests that there is either gradual and gentle change, or else abrupt and violent change. In fact, all grades between these two extremes, and combinations of gentle and violent processes, are conceivable. It seems reasonable to suggest that land-surface history has involved a combination of gentle and violent processes.

Modelling in geomorphology

In trying to single out the components and interrelations of geomorphic systems, some degree of abstraction or simplification is necessary: the landscape is too rich a mix of objects and interactions to account for all components and relationships in them. The process of simplifying real landscapes to manageable proportions is **model building**. Defined in a general way, a geomorphic **model** is a simplified representation of some aspect of a real landscape that happens to interest a geomorphologist. It is an attempt to describe, analyse, simplify, or display a geomorphic system (cf. Strahler 1980).

Geomorphologists, like all scientists, build models at different levels of abstraction (Figure 1.12). The simplest level involves a change of scale. In this

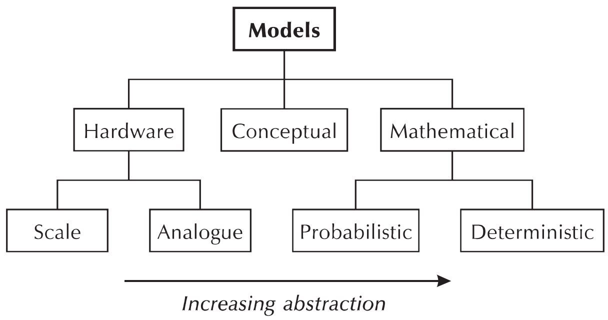


Figure 1.12 Types of model in geomorphology.
Source: After Huggett (1993, 4)

case, a **hardware model** represents the system (see Mosley and Zimpfer 1978). There are two chief kinds of hardware model: scale models and analogue models. **Scale** (or **iconic**) **models** are miniature, or sometimes gigantic, copies of systems. They differ from the systems they represent only in size. Relief models, fashioned out of a suitable material such as plaster of Paris, have been used to represent topography as a three-dimensional surface. Scale models need not be static: models made using materials identical to those found in Nature, but with the dimensions of the system scaled down, can be used to simulate dynamic behaviour. In practice, scale models of this kind imitate a portion of the real world so closely that they are, in effect, 'controlled' natural systems. An example is Stanley A. Schumm's (1956) use of the badlands at Perth Amboy, New Jersey, to study the evolution of slopes and drainage basins. The great advantage of this type of scale model, in which the geometry and dynamics of the model and system are virtually identical, is that the investigator wields a high degree of control over the simplified experimental conditions. Other scale models use natural materials, but the geometry of the model is dissimilar to the geometry of the system it imitates – the investigator scales down the size of the system. The process of reducing the size of a system creates a number of awkward problems associated with scaling. For instance, a model of the Severn Estuary made at a scale of 1 : 10,000 can easily preserve geometrical and topographical relationships. However, when adding water, an actual depth of water of, say, 7 m is represented in the model by a layer of water less than 0.7 mm deep. In such a thin layer of water, surface tensions will cause enormous problems, and it will be impossible to simulate tidal range and currents. Equally, material scaled down to represent sand in the real system would be so tiny that most of it would float. These problems of scaling are usually surmountable, to a certain extent at least, and scale models are used to mimic the behaviour of a variety of geomorphic systems. For example, scale models have assisted studies of the dynamics of rivers and river systems using waterproof troughs and flumes.

Analogue models are more abstract scale models. The most commonly used analogue models are maps and remotely sensed images. On a map, the surface features of a landscape are reduced in scale and represented by symbols: rivers by lines, relief by contours, and spot heights by points, for instance. Remotely sensed images represent, at a reduced scale, certain properties of the landscape systems. Maps and remotely sensed images are, except where a series of them be available for different times, static analogue models. Dynamic analogue models may also be built. They are hardware models in which the system size is changed, and in which the materials used are analogous to, but not the same as, the natural materials of the system. The analogous materials simulate the dynamics of the real system. In a laboratory, the clay kaolin can be used in place of ice to model the behaviour of a valley glacier. Under carefully controlled conditions, many features of valley glaciers, including crevasses and step faults, develop in the clay. Difficulties arise in this kind of analogue model, not the least of which is the problem of finding a material that has mechanical properties comparable to the material in the natural system.

Conceptual models are initial attempts to clarify loose thoughts about the structure and function of a geomorphic system. They often form the basis for the construction of mathematical models. **Mathematical models** translate the ideas encapsulated in a conceptual model into the formal, symbolic logic of mathematics. The language of mathematics offers a powerful tool of investigation limited only by the creativity of the human mind. Of all modes of argument, mathematics is the most rigorous. Nonetheless, the act of quantification, of translating ideas and observations into symbols and numbers, is in itself nothing unless validated by explanation and prediction. The art and science of using mathematics to study geomorphic systems is to discover expressions with explanatory and predictive powers. These powers set mathematical models apart from conceptual models. An unquantified conceptual model is not susceptible of formal proof; it is simply a body of ideas. A mathematical model, on the other hand, is testable by matching predictions against the yardstick of observation. By a continual process of mathematical model building, model testing, and model redesign, the understanding of the form and function of geomorphic systems should advance.

Three chief classes of mathematical model assist the study of geomorphic systems: stochastic models, statistical models, and deterministic models. The first two classes are both probabilistic models. **Stochastic models** have a random component built into them that describes a system, or some facet of it, based on probability. **Statistical models**, like stochastic models, have random components. In statistical models, the random components represent unpredictable fluctuations in laboratory or field data that may arise from measurement error, equation error, or the inherent variability of the objects being measured. A body of inferential statistical theory exists that determines the manner in which the data should be collected and how relationships between the data should be managed. Statistical models are, in a sense, second best to deductive models: they can be applied only under strictly controlled conditions, suffer from a number of deficiencies, and are perhaps most profitably employed only when the 'laws' determining system form and process are poorly understood. **Deterministic models** are conceptual models expressed mathematically and containing no random components. They are derivable from physical and chemical principles without recourse to experiment. It is sound practice, therefore, to test the validity of a deterministic model by comparing its predictions with independent observations made in the field or the laboratory. Hillslope models based on the conservation of mass are examples of deterministic models (p. 175).

HISTORY

Historical geomorphologists study landform evolution or changes in landforms over medium and long timescales, well beyond the span of an individual human's experience – centuries, millennia, millions and hundreds of millions of years. Such considerations go well beyond the short-term predictions of the process modellers. They bring in the historical dimension of the subject with all its attendant assumptions and methods. Historical geomorphology relies mainly on the form of the land surface and on the sedimentary record for its databases.

Reconstructing geomorphic history

The problem with measuring geomorphic processes is that, although it establishes current operative processes and their rates, it does not provide a dependable guide to processes that were in action a million years ago, ten thousand years ago, or even a hundred years ago. Some landform features may be inherited from the past and are not currently forming. In upland Britain, for instance, hillslopes sometimes bear ridges and channels that were fashioned by ice and meltwater during the last ice age. In trying to work out the long-term evolution of landforms and landscapes, geomorphologists have three options open to them – modelling, chronosequence studies, and stratigraphic reconstruction.

Mathematical models of the hillslopes predict what happens if a particular combination of slope processes is allowed to run on a hillslope for millions of years, given assumptions about the initial shape of the hillslope, tectonic uplift, tectonic subsidence, and conditions at the slope base (the presence or absence of a river, lake, or sea). Some geomorphologists would argue that these models are of limited worth because environmental conditions will not stay constant, or even approximately constant, for that long. Nonetheless, the models do show the broad patterns of hillslope and land-surface change that occur under particular process regimes. Some examples of long-term hillslope models will be given in Chapter 7.

Stratigraphic and environmental reconstruction

Fortunately for researchers into past landscapes, several archives of past environmental conditions exist: tree rings, lake sediments, polar ice cores, mid-latitude ice cores, coral deposits, loess, ocean cores, pollen, palaeosols, sedimentary rocks, and historical records (see Huggett 1997b, 8–21). Sedimentary deposits are an especially valuable source of information about past landscapes. In some cases, geomorphologists may apply the **principles of stratigraphy** to the deposits to establish a relative sequence of events. Colluvium for example, which builds up towards a hillslope base, is commonly deposited episodically. The result is that distinct layers are evident in a section, the upper layers being progressively younger than the lower layers. If such techniques as **radiocarbon dating** or **dendrochronology** can date these sediments, then they may provide an absolute timescale for the past activities on the hillslope, or at least the past activities that have left traces in the sedimentary record. Recognizing the origin of the deposits may also be possible – glacial, periglacial, colluvial, or whatever. And sometimes geomorphologists use techniques of environmental reconstruction to establish the climatic and other environmental conditions at the time of sediment deposition.

The recent global environmental change agenda has given environmental reconstruction techniques a fillip. A core project of the IGBP (International Geosphere–Biosphere Programme) is called **Past Global Changes** (**PAGES**). It concentrates on two slices of time: (1) the last 2,000 years of Earth history, with a temporal resolution of decades, years, and even months; and (2) the last several hundred thousand years, covering glacial–interglacial cycles, in the hope of providing insights into the processes that induce global change (IGBP 1990). Examples of geomorphological contributions to environmental change over these timescales may be found in the book *Geomorphology, Human Activity and Global Environmental Change* edited by Olav Slaymaker (2000a).

Landform chronosequences

Another option open to the historical geomorphologist is to find a site where a set of landforms differ from place to place and where that spatial sequence of landforms may be interpreted as a time sequence. Such sequences are called **topographic chronosequences**, and the procedure is sometimes referred to as **space–time substitution** or, using a term borrowed from physics, **ergodicity**.

Charles Darwin used the chronosequence method to test his ideas on coral-reef formation. He thought that barrier reefs, fringing reefs, and atolls occurring at different places represented different evolutionary stages of island development applicable to any subsiding volcanic peak in tropical waters. William Morris Davis applied this evolutionary schema to landforms in different places and derived what he deemed was a time sequence of landform development – the geographical cycle – running from youth, through maturity, to senility. This seductively simple approach is open to misuse. The temptation is to fit the landforms into some preconceived view of landscape change, even though other sequences might be constructed. A study of south-west African landforms since Mesozoic times highlights the significance of this problem, where several styles of landscape evolution were consistent with the observed history of the region (Gilchrist *et al.* 1994). Users of the method must also be warned that not all spatial differences are temporal differences – factors other than time exert a strong influence on the form of the land surface, and landforms of the same age might differ through historical accidents. Moreover, it pays to be aware of **equifinality**, the idea that different sets of processes may produce the same landform. The converse of this idea is that land form is an unreliable guide to process. Given these consequential difficulties, it is best to treat chronosequences circumspectly.

Trustworthy topographic chronosequences are rare. The best examples normally come from artificial landscapes, though there are some landscapes in which, by quirks of history, spatial differences are translatable into time sequences. Occasionally, field conditions lead to adjacent hillslopes being progressively removed from the action of a fluvial or marine process at their bases. This has happened along a segment of the South Wales coast, in the British Isles, where cliffs are formed in Old Red Sandstone (Savigear 1952, 1956). Originally, the coast between Gilman Point and the Taff estuary was exposed to wave action. A sand spit started to grow. Wind-blown and marsh deposits accumulated between the spit and the original shoreline, causing the sea progressively to abandon the cliff base from west to east. The present cliffs are thus a topographic chronosequence: the cliffs furthest west have been subject to subaerial

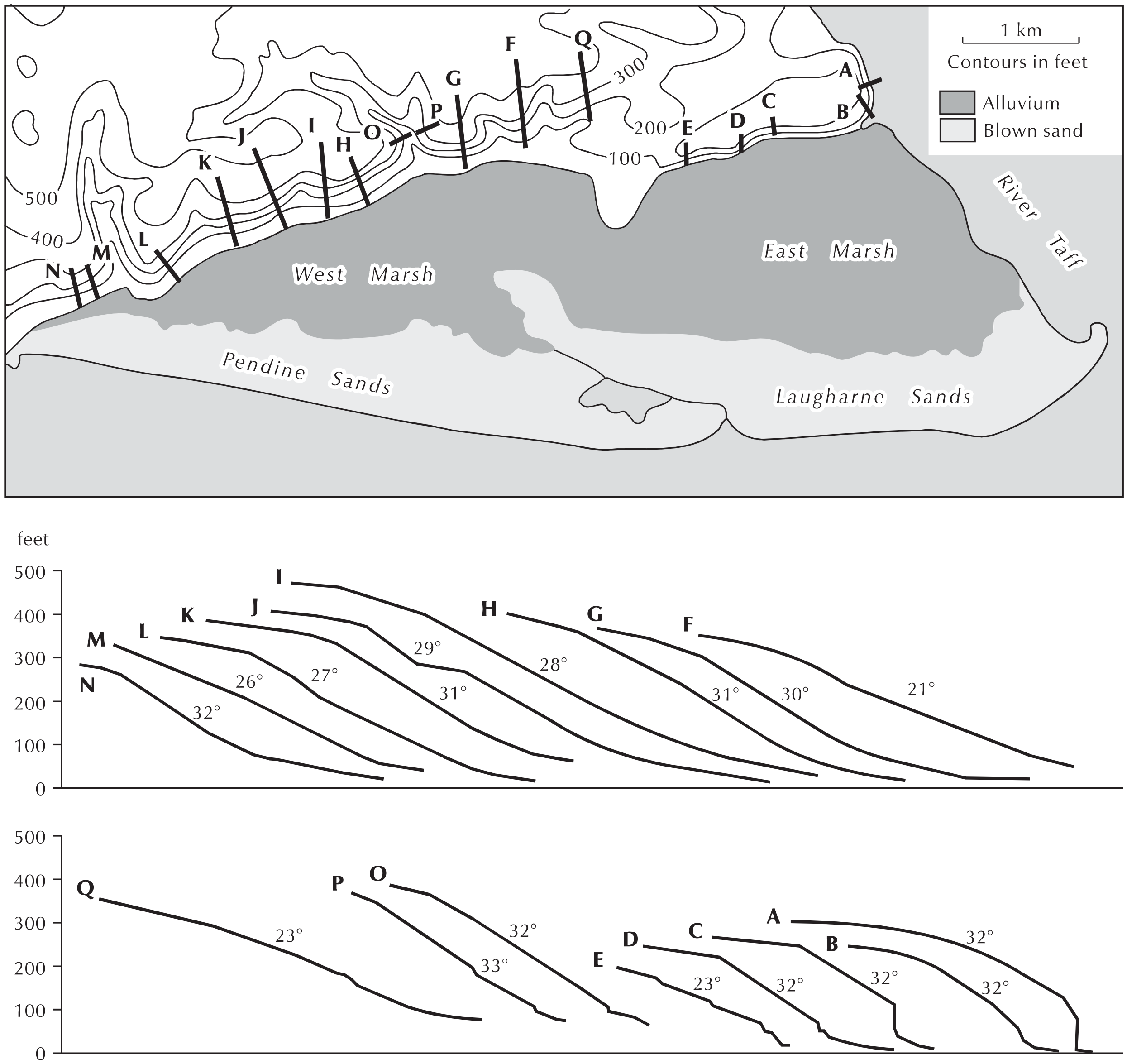


Figure 1.13 A topographic chronosequence in South Wales. (a) The coast between Gilman Point and the Taff estuary. The sand spit has grown progressively from west to east so that the cliffs to the west have been longest-protected from wave action. (b) The general form of the hillslope profiles located on Figure 1.11a. Cliff profiles become progressively older in alphabetical order, A–N.

Source: From Huggett (1997, 238) after Savigear (1952, 1956) denudation without waves cutting their base the longest, while those to the east are progressively younger (Figure 1.13). Slope profiles along Port Hudson bluff, on the Mississippi River in Louisiana, southern USA, reveal a chronosequence (Brunsden and Kesel 1973). The Mississippi River was undercutting the entire bluff segment in 1722. Since then, the channel has shifted about 3 km downstream with a concomitant cessation of undercutting. The changing conditions at the slope bases have reduced the mean slope angle from 40° to 22°.

The question of scale

A big problem faced by geomorphologists is that, as the size of geomorphic systems increases, the explanations of their behaviour may change. Take the case of a fluvial system. The form and function of a larger-scale drainage network require a different explanation from a smaller-scale meandering river within the network, and an even smaller-scale point bar along the meander requires a different explanation again. The process could carry on down through bedforms on the point bar, to the position and nature of individual sediment grains within the bedforms (cf. Schumm 1985a; 1991, 49). A similar problem applies to the time dimension. Geomorphic systems may be studied in action today. Such studies are short-term, lasting for a few years or decades. Yet geomorphic systems have a history that goes back centuries, millennia, or millions of years. Using the results of short-term studies to predict how geomorphic systems will change over long periods is difficult. Stanley A. Schumm (1985, 1991) tried to resolve the **scale problem**, and in doing so established some links between process and historical studies. He argued that, as the size and age of a landform increase, so present conditions can explain fewer of its properties and geomorphologists must infer more about its past. Figure 1.14 summarizes his idea. Evidently, such small-scale landforms and processes as sediment movement and river bedforms may be understood with recent historical information. River channel morphology may have a considerable historical component, as when rivers flow on alluvial plain surfaces that events during the Pleistocene determined. Large-scale landforms, such as structurally controlled drainage networks and mountain ranges, are explained mainly by historical information. A corollary of this idea is that the older and bigger a landform, the less accurate will be predictions and postdictions about it based upon present conditions. It also shows that an understanding of landforms requires a variable mix of process geomorphology and historical geomorphology, and that the two subjects should work together rather than stand in polar opposition.

UNIFORMITY AND NON-UNIFORMITY: A NOTE ON METHODOLOGY

Process and historical geomorphologists alike face a problem with their methodological base. In practising their trade, all scientists, including geomorphologists, follow rules. Scientific practitioners established these rules, or guidelines. They advise scientists how to go about the business of making scientific enquiries. In other words, they are guidelines concerned with scientific methodology or procedures. The foremost guideline – the **uniformity of law** – is the premise from which all scientists work. It is the presupposition that natural laws are invariant in time and space. In simple terms, this means that, throughout Earth history, the laws of physics, chemistry, and biology have always been the same. Water has always flowed downhill, carbon dioxide has always been a greenhouse gas, and most living things have always depended upon carbon, hydrogen, and oxygen. Three other guidelines are relevant to geomorphology. Unlike the uniformity of law, which is a universally accepted basis for scientific investigation, they are substantial claims or suppositions about how the Earth works and are open to interpretation. First, the **principle of simplicity** or, as it is commonly called in geomorphology, the **uniformity of process** states that no extra, fanciful, or unknown causes should be invoked if available processes will do the job. It is the supposition of **actualism**, the belief that past events are the outcome of processes seen in operation today. However, the dogma of actualism is being challenged, and its flip-side – **non-actualism** – is gaining ground. Some geologists and geomorphologists are coming round to the view that the circumstances under which processes acted in the past were very different from those experienced today, and

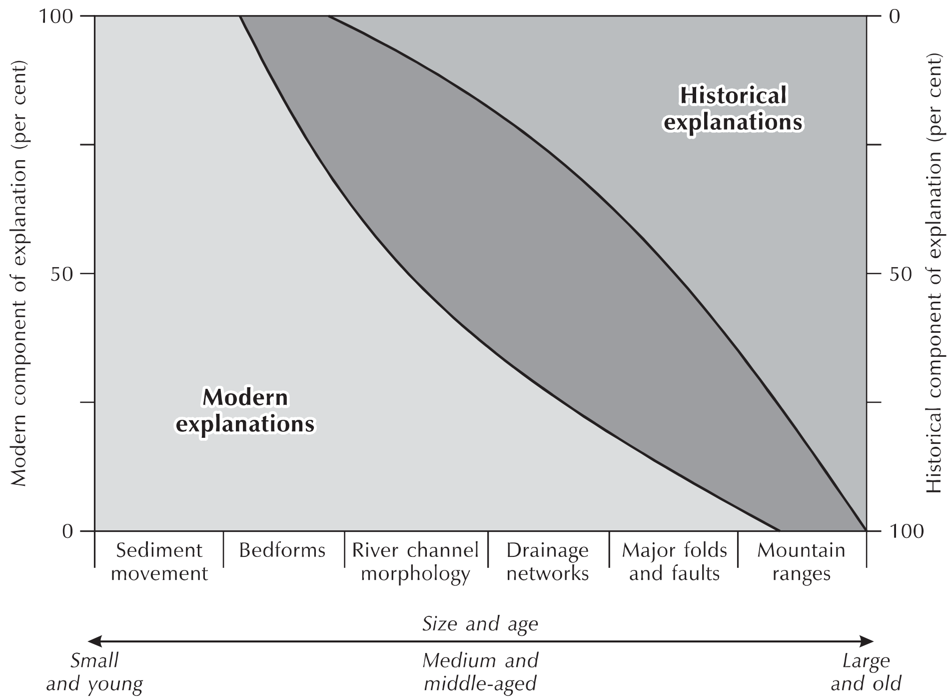


Figure 1.14 The components of historical explanation needed to account for geomorphic events of increasing size and age. The top right of the diagram contains purely historical explanations, while the bottom left contains purely modern explanations. The two explanations overlap in the middle zone, the top curve showing the maximum extent of modern explanations and the lower curve showing the maximum extent of historical explanations.
Source: After Schumm (1985b, 1991, 53)

that those differences greatly influence the interpretation of past processes. So, before the evolution of land plants, and especially the grasses, the processes of weathering, erosion, and deposition would have occurred in a different context, and Palaeozoic deserts, or even Permian deserts, may not directly correspond to modern deserts. The second substantive claim concerns the rate of Earth surface processes, two extreme views being **gradualism** and **catastrophism** (p. 21). The third substantive claim concerns the changing state of the Earth's surface, **steady-statism** arguing for a more or less constant state, or at least cyclical changes about a comparatively invariant mean state, and **directionalism** arguing in favour of directional changes.

Uniformitarianism is a widely used, but too often loosely used, term in geomorphology. A common mistake is to equate uniformitarianism with actualism. Uniformitarianism was a system of assumptions about Earth history argued by Charles Lyell, the nineteenth-century geologist. Lyell articulately advocated three 'uniformities', as well as the uniformity of law: the uniformity of process (actualism), the uniformity of rate (gradualism), and the uniformity of state (steady-statism). Plainly, extended to geomorphology, uniformitarianism, as introduced by Lyell, is a set of beliefs about Earth surface processes and states. Other sets of beliefs are possible. The diametric opposite of Lyell's uniformitarian position would be a belief in the non-uniformity of process (non-actualism), the non-uniformity of rate (catastrophism), and the non-uniformity of state (directionalism). All other combinations of assumption are possible and give rise to different 'systems of Earth history' (Huggett 1997a). The various systems may be tested against field evidence. To be sure, directionalism was accepted even before Lyell's death, and non-actualism and, in particular, catastrophism are discussed in geomorphological circles.

SUMMARY

Geomorphology is the study of landforms. Three key elements of geomorphology are land form, geomorphic process, and land-surface history. The three main brands of geomorphology are process (or functional) geomorphology, applied geomorphology, and historical geomorphology. Other brands include tectonic geomorphology, submarine geomorphology, planetary geomorphology, and climatic geomorphology. Form is described by morphological maps or, more recently, by geomorphometry. Geomorphometry today uses digital elevation models and is a sophisticated discipline. Armed with a powerful combination of predictive models, field observations, and laboratory experiments, process geomorphologists study geomorphic processes in depth. They commonly use a systems approach to their subject. Form systems, flow or cascading systems, and process–form or process–response systems are all recognized. Negative feedback and positive feedback relationships are significant features in the dynamics of geomorphic systems. The great achievements of process geomorphology include notions of stability, instability, and thresholds in landscapes, the last two of which belie simplistic ideas on cause and effect in landscape evolution. Uncertainty surrounds the issue of geomorphic process rates. Magnitude and frequency impinge on part of this uncertainty. At first it was believed that medium-magnitude and medium-frequency events did the greatest geomorphic work. Some studies now suggest that rare events such as immense floods may have long-lasting effects on landforms. Land-surface history is the domain of the historical geomorphologist. Some early historical work was criticized for reading too much into purely morphological evidence. Nonetheless, historical geomorphologists had some great successes by combining careful field observation with the analysis of the sedimentary record. Historical geomorphologists reconstruct past changes in landscapes using the methods of environmental and stratigraphic reconstruction or topographic chronosequences, often hand in hand with dating techniques. Geomorphology has engaged in methodological debates over the extent to which the present is a key to the past and the rates of Earth surface processes.

ESSAY QUESTIONS

1. **To what extent do process geomorphology and historical geomorphology inform each other?**
2. **Discuss the pros and cons of a 'systems approach' in geomorphology.**
3. **Explain the different types of equilibrium and non-equilibrium recognized in geomorphic systems.**

FURTHER READING

Ahnert, F. (1998) *Introduction to Geomorphology*. London: Arnold.
A good starting text with many unusual examples.

Bloom, A. L. (1998) *Geomorphology: A Systematic Analysis of Late Cenozoic Landforms*, 3rd edn. Upper Saddle River, N. J., and London: Prentice Hall.
A sound text with a focus on North America.

Goudie, A. S. (ed.) (1994) *Geomorphological Techniques*, 2nd edn. London and New York: Routledge.
Covers the topics not covered by the present book – how geomorphologists measure form and process.

Kennedy, B. A. (2005) *Inventing the Earth: Ideas on Landscape Development Since 1740*. Oxford: Blackwell.
A good read on the relatively recent history of ideas about landscape development.

Ritter, D. F., Kochel, R. C., and Miller, J. R. (1995) *Process Geomorphology*, 3rd edn. Dubuque, Ill., and London: William C. Brown.
A good, well-illustrated, basic text with a fondness for North American examples.

Strahler, A. H. and Strahler, A. N. (2006) *Introducing Physical Geography*, 4th edn. New York: John Wiley & Sons.
Comprehensive and accessible coverage of all aspects of physical geography if a general background is needed.

Summerfield, M. A. (1991) *Global Geomorphology: An Introduction to the Study of Landforms*. Harlow, Essex: Longman.
A classic after just fifteen years. Includes material on the geomorphology of other planets.

Thorn, C. E. (1988) *An Introduction to Theoretical Geomorphology*. Boston, Mass.: Unwin Hyman.
A very clear discussion of the big theoretical issues in geomorphology. Well worth a look.

2

THE GEOMORPHIC SYSTEM

The Earth's topography results from the interplay of many processes, some originating inside the Earth, some outside it, and some on it. This chapter covers:

- grand cycles of water and rock
- the wearing away and the building up of the land surface
- tectonics, erosion, and climate

The Earth's surface in action: mountain uplift and global cooling

Over the last 40 million years, the uplift of mountains has been a very active process. During that time, the Tibetan Plateau has risen by up to 4,000 m, with at least 2,000 m in the last 10 million years. Two-thirds of the uplift of the Sierra Nevada in the USA has occurred in the past 10 million years. Similar changes have taken place (and are still taking place) in other mountainous areas of the North American west, in the Bolivian Andes, and in the New Zealand Alps. This period of active mountain building seems to be linked to global climatic change, in part through airflow modification and in part through weathering. Young mountains weather and erode quickly. Weathering processes remove carbon dioxide from the atmosphere by converting it to soluble carbonates. The carbonates are carried to the oceans, where they are deposited and buried. It is possible that the growth of the Himalaya scrubbed enough carbon dioxide from the atmosphere to cause a global climatic cooling that culminated in the Quaternary ice ages (Raymo and Ruddiman 1992; Ruddiman 1997). This shows how important the geomorphic system can be to environmental change.

ROCK AND WATER CYCLES

The Earth's surface – the **toposphere** – sits at the interfaces of the solid lithosphere, the gaseous atmosphere, and the watery hydrosphere. It is also the dwelling-place of many living things. Gases, liquids, and solids are exchanged between these spheres in three grand cycles, two of which – the **water** or **hydrological cycle** and the **rock cycle** – are crucial to understanding landform evolution. The third grand cycle – the **biogeochemical cycle** – is the circulation of chemical elements (carbon, oxygen, sodium, calcium, and so on) through the upper mantle, crust, and ecosphere, but is less significant to landform development, although some biogeochemical cycles regulate the composition of the atmosphere, which in turn can affect weathering.

Water cycle

The **hydrosphere** – the surface and near-surface waters of the Earth – is made of **meteoric water**. The water cycle is the circulation of meteoric water through the hydrosphere, atmosphere, and upper parts of the crust. It is linked to the circulation of deep-seated **juvenile water** associated with magma production and the rock cycle. Juvenile water ascends from deep rock layers through volcanoes, where it issues into the meteoric zone for the first time. On the other hand, meteoric water held in hydrous minerals and pore spaces in sediments, known as connate water, may be removed from the meteoric cycle at subduction sites, where it is carried deep inside the Earth.

The land phase of the water cycle is of special interest to geomorphologists. It sees water transferred from the atmosphere to the land and then from the land back to the atmosphere and to the sea. It includes a **surface drainage system** and a **subsurface drainage system**. Water flowing within these drainage systems tends to be organized within **drainage basins**, which are also called **watersheds** in the USA and **catchments** in the UK. The basin water system may be viewed as a set of water stores that receive inputs from the atmosphere and deep inflow from deep groundwater storage, that lose outputs through evaporation and streamflow and deep outflow, and that are linked by internal flows. In summary, the basin water runs like this. Precipitation entering the system is stored on the soil or rock surface, or is intercepted by vegetation and stored there, or falls directly into a stream channel. From the vegetation it runs down branches and trunks (stemflow), or drips off leaves and branches (leaf and stem drip), or it is evaporated. From the soil or rock surface, it flows over the surface (overland flow), infiltrates the soil or rock, or evaporates. Once in the rock or soil, water may move laterally down hillsides (throughflow, pipeflow, interflow) to feed rivers, or it may move downwards to recharge groundwater storage, or it may evaporate. Groundwater may rise by capillary action to top up the rock and soil water stores, or it may flow into a stream (baseflow), or may exchange water with deep storage.

Rock cycle

The **rock cycle** is the repeated creation and destruction of crustal material – rocks and minerals (Box 2.1). Volcanoes, folding, faulting, and uplift all bring igneous and other rocks, water, and gases to the base of the atmosphere and hydrosphere. Once exposed to the air and meteoric water, these rocks begin to decompose and disintegrate by the action of weathering. Gravity, wind, and water transport the weathering products to the oceans. Deposition occurs on the ocean floor. Burial of the loose sediments leads to compaction, cementation, and recrystallization, and so to the formation of sedimentary rocks. Deep burial may convert sedimentary rocks into metamorphic rocks. Other deep-seated processes may produce granite. If uplifted, intruded or extruded, and exposed at the land surface, the loose sediments, consolidated sediments, metamorphic rocks, and granite may join in the next round of the rock cycle.

Volcanic action, folding, faulting, and uplift may all impart potential energy to the toposphere, creating the 'raw relief' on which geomorphic agents may act to fashion the marvellously multifarious array of landforms found on the Earth's surface – the physical toposphere. Geomorphic or exogenic agents are wind, water, waves, and ice, which act from outside or above the toposphere; these contrast with endogenic (tectonic and volcanic) agents, which act upon the toposphere from inside the planet.

Box 2.1

ROCKS AND MINERALS

The average composition by weight of chemical **elements** in the lithosphere is oxygen 47 per cent, silicon 28 per cent, aluminium 8.1 per cent, iron 5.0 per cent, calcium 3.6 per cent, sodium 2.8 per cent, potassium 2.6 per cent, magnesium 2.1 per cent, and the remaining eighty-three elements 0.8 per cent. These elements combine to form **minerals**. The chief minerals in the lithosphere are **feldspars** (aluminium silicates with potassium, sodium, or calcium), **quartz** (a form of silicon dioxide), **clay minerals** (complex aluminium silicates), **iron minerals** such as limonite and hematite, and **ferromagnesian minerals** (complex iron, magnesium, and calcium silicates). **Ore deposits** consist of common minerals precipitated from hot fluids. They include pyrite (iron sulphide), galena (lead sulphide), blende or sphalerite (zinc sulphide), and cinnabar (mercury sulphide).

Rocks are mixtures of crystalline forms of minerals. There are three main types: igneous, sedimentary, and metamorphic.

Igneous rocks

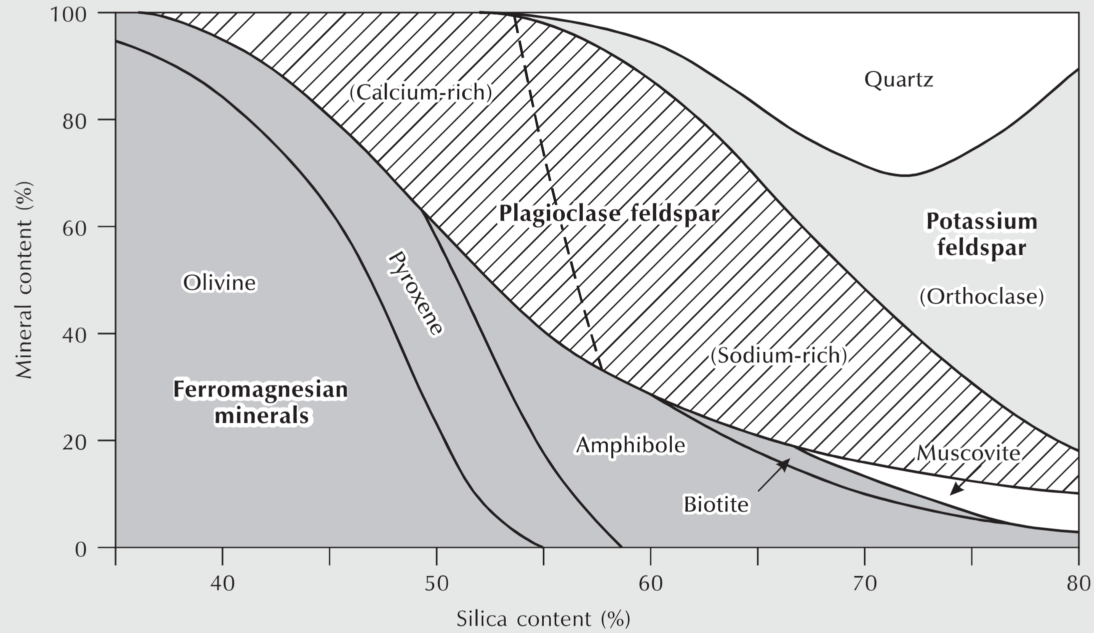
These form by solidification of molten rock (magma). They have varied compositions (Figure 2.1). Most igneous rocks consist of **silicate minerals**, especially those of the felsic mineral group, which comprises quartz and feldspars (potash and plagioclase). **Felsic minerals** have silicon, aluminium, potassium, calcium, and sodium as the dominant elements. Other important mineral groups are the **micas**, **amphiboles**, and **pyroxenes**. All three groups contain aluminium, magnesium, iron, and potassium or calcium as major elements. Olivine is a magnesium and iron silicate. The micas, amphiboles (mainly hornblende), pyroxenes, and olivine constitute the **mafic minerals**, which are darker in colour and denser than the felsic minerals. **Felsic rocks** include diorite, tonalite, granodiorite, rhyolite, andesite, dacite, and granite. **Mafic rocks** include gabbro and basalt. **Ultramafic rocks**, which are denser still than mafic rocks, include peridotite and serpentine. Much of the lithosphere below the crust is made of peridotite. Eclogite is an ultramafic rock that forms deep in the crust, nodules of which are sometimes carried to the surface by volcanic action. At about 400 km below the surface, olivine undergoes a phase change (it fits into a more tightly packed crystal lattice whilst keeping the same chemical composition) to spinel, a denser silicate mineral. In turn, at about 670 km depth, spinel undergoes a phase change into perovskite, which is probably the chief mantle constituent and the most abundant mineral in the Earth.

Sedimentary rocks

These are layered accumulations of mineral particles derived mostly from weathering and erosion of pre-existing rocks. They are clastic, organic, or chemical in origin. **Clastic sedimentary rocks** are unconsolidated or indurated sediments (boulders, gravel, sand, silt, clay) derived from geomorphic processes. Conglomerate, breccia, sandstone, mudstone, claystone, and shale are examples. **Organic sedimentary rocks** and **mineral fuels** form from organic materials. Examples are coal, petroleum, and natural gas. **Chemical sedimentary rocks** form by chemical precipitation in oceans, seas, lakes, caves, and, less commonly, rivers. Limestone, dolomite, chert, tufa, and evaporites are examples.

Metamorphic rocks

These form through physical and chemical changes in igneous and sedimentary rocks. Temperatures or pressures high enough to bring about recrystallization of the component minerals cause the changes. Slate, schist, quartzite, marble, and gneiss are examples.



	ULTRABASIC	BASIC	INTERMEDIATE		ACID
EXTRUSIVE Fine-grained	(rare)	Basalt	Andesite	Trachyte	Rhyolite Obsidian
INTRUSIVE Medium-grained	(rare)	Dolerite	Microdiorite	Microsyenite	Microgranite
INTRUSIVE Coarse-grained	Peridotite	Gabbro	Diorite	Syenite	Granite

Figure 2.1 Igneous rocks and their component minerals. The classification is based on the silica content, which produces an ultrabasic–acid axis. The terms ‘acid’ and ‘basic’ are not meant to suggest that the rocks are acidic or alkaline in the customary sense, but merely describe their silica content.

The surface phase, and particularly the land-surface phase, of the rock cycle is the domain of geomorphologists. The flux of materials across the land surface is, overall, unidirectional and is a cascade rather than a cycle. The basics of the **land-surface debris cascade** are as follows. Weathering agents move into the soil and rock along a weathering front, and in doing so fresh rock is brought into the system. Material may be added to the land surface by deposition, having been borne by wind, water, ice, or animals. All the materials in the system are subject to transformations by the complex processes of weathering. Some weathering products revert to a rock-like state by further transformations: under the right conditions, some chemicals precipitate out from solution to form hardpans and crusts. And many organisms produce resistant organic and inorganic materials to shield or to support their bodies. The weathered mantle may remain in place or it may move downhill. It may creep, slide, slump, or flow downhill under the influence of gravity (mass movements), or moving water may be wash or carry it downhill. In addition, the wind may erode it and take it elsewhere.

The land-surface debris cascade produces landforms. It does so partly by selectively weathering and eroding weaker rocks (Box 2.2).

Biogeochemical cycles

The **biosphere** powers a global cycle of carbon, oxygen, hydrogen, nitrogen, and other mineral elements. These minerals circulate with the ecosphere and are exchanged between the ecosphere and its environment. The circulations are called **biogeochemical cycles**. The land phase of these cycles is intimately linked with water and debris movements.

Interacting cycles

The water cycle and the rock cycle interact (Figure 2.2). John Playfair was perhaps the first person to recognize this crucial interaction in the Earth system, and he was perhaps the great-grandfather of **Earth System**

Box 2.2

ROCKS AND RELIEF

The ability of rocks to resist the agents of denudation depends upon such factors as particle size, hardness, porosity, permeability, the degree to which particles are cemented, and mineralogy. **Particle size** determines the surface area exposed to chemical attack: gravels and sands weather slowly compared with silts and clays. The hardness, mineralogy, and degree of rock cementation influences the rate at which weathering decomposes and disintegrates them: a siliceous sandstone is more resistant to weathering than a calcareous sandstone. **Permeability** is an important property in shaping weathering because it determines the rate at which water seeps into a rock body and dictates the internal surface area exposed to weathering (Table 2.1).

As a rule, igneous and metamorphic rocks are resistant to weathering and erosion. They tend to form the basements of cratons, but where they are exposed at the surface or are thrust through the overlying sedimentary cover by tectonic movements they often give rise to resistant hills. English examples are the Malvern Hills in Hereford and Worcester, which have a long and narrow core of gneisses, and Charnwood Forest in the Midlands, which is formed of Precambrian volcanic and plutonic rocks. The strongest igneous and metamorphic rocks are quartzite, dolerite, gabbro, and basalt, followed by marble, granite, and gneiss. These resistant rocks tend to form relief features in landscapes. The quartz-dolerite Whin Sill of northern England is in places is a prominent topographic feature (p. 119). Basalt may cap plateaux and other sedimentary hill features. Slate is a moderately strong rock, while schist is weak.

Sedimentary rocks vary greatly in their ability to resist weathering and erosion. The weakest of them are chalk and rock salt. However, the permeability of chalk compensates for its weakness and chalk resists denudation, sometimes with the help of more resistant bands within it, to form cuestas (p. 133), as in the North and South Downs of south-east England. Coal, claystone, and siltstone are weak rocks that offer little resistance to erosion and tend to form vales. An example from south-east England is the lowland developed on the thick Weald Clay. Sandstone is a moderately strong rock that may form scarps and cliffs. Whether or not it does so depends upon the nature of the sandstone and the environment in which it is found (e.g. Robinson and Williams 1994). Clay-rich or silty sandstones are often cemented weakly, and the clay reduces their permeability. In temperate European environments, they weather and are eroded readily and form low relief, as is the case with the Sandgate Beds of the Lower Greenland, south-east England. In arid regions, they may produce prominent cuestas. Weakly cemented sands and sandstones that contain larger amounts of quartz often form higher ground in temperate Europe, probably because their greater porosity reduces runoff

Table 2.1 Porosities and permeabilities of rocks and sediments

Material	*Representative porosity (per cent void space)*	*Permeability range (litres/day/m²)*
Unconsolidated		
Clay	50–60	0.0004–0.04
Silt and glacial till	20–40	0.04–400
Alluvial sands	30–40	400–400,000
Alluvial gravels	25–35	400,000–40,000,000
Indurated: sedimentary		
Shale	5–15	0.000004–0.004
Siltstone	5–20	0.0004–40
Sandstone and conglomerate	5–25	0.04–4,000
Limestone	0.1–10	0.004–400
Indurated: igneous and metamorphic		
Volcanic (basalt)	0.001–50	0.004–40
Granite (weathered)	0.001–10	0.0004–0.4
Granite (fresh)	0.0001–1	0.000004–0.0004
Slate	0.001–1	0.000004–0.004
Schist	0.001–1	0.00004–0.04
Gneiss	0.0001–1	0.000004–0.004
Tuff	10–80	0.0004–40

Source: Adapted from Waltz (1969)

and erosion. A case in point is the Folkestone Sands of south-east England, which form a low relief feature in the northern and western margins of the Weald, though it is overshadowed by the impressive Hythe Beds cuesta. Interestingly, the Hythe Beds comprise incoherent sands over much of the Weald, but in the west and north-west they contain sandstones and chert beds, and in the north and north-east the sands are partly replaced by interbedded sandy limestones and loosely cemented sandstones. These resistant bands produce a discontinuous cuesta that is absent in the south-eastern Weald, but elsewhere rises to form splendid ramparts at Hindhead (273 m), Blackdown (280 m), and Leith Hill (294 m) that tower above the Low Weald (Jones 1981, 18). However, in general, hillslopes on the aforementioned sandstones are rarely steep and usually covered with soil. Massive and more strongly cemented sandstones and gritstones normally form steep slopes and commonly bear steep cliffs and isolated pillars. They do so throughout the world.

Details of the influence of rocks upon relief will be discussed in Chapters 4 and 5.

Science (Box 2.3). Here is how he described it in old-fashioned but most elegant language:

> We have long been accustomed to admire that beautiful contrivance in Nature, by which the water of the ocean, drawn up in vapour by the atmosphere, imparts in its descent, fertility to the earth, and becomes the great cause of vegetation and of life; but now we find, that this vapour not only fertilizes, but creates the soil; prepares it from the soil rock, and, after employing it in the great operations of the surface, carries it back into the

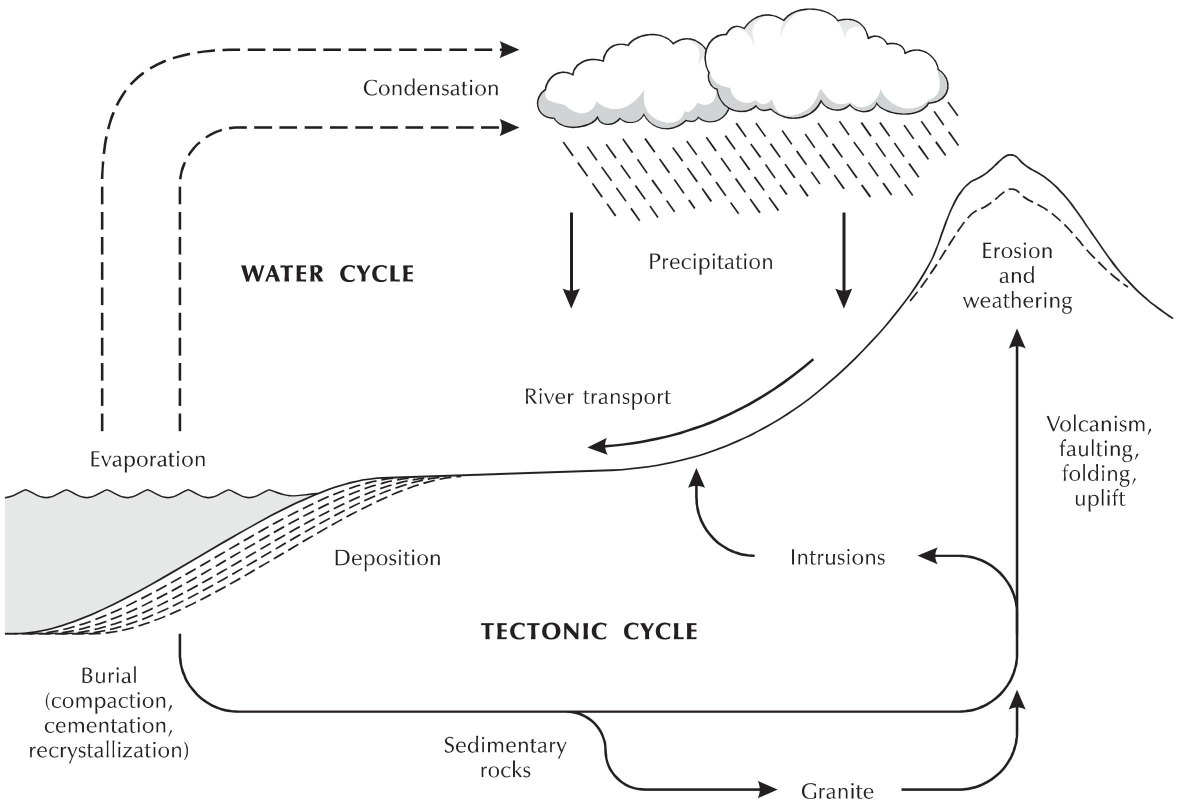


Figure 2.2 The rock cycle, the water cycle, and their interaction.

> regions where all its mineral characters are renewed. Thus, the circulation of moisture through the air, is a prime mover, not only in the annual succession of seasons, but in the great geological cycle, by which the waste and reproduction of entire continents is circumscribed.
>
> (Playfair 1802, 128)

DENUDATION AND DEPOSITION

Weathering is the decay of rocks by biological, chemical, and mechanical agents with little or no transport. It produces a mantle of rock waste. The **weathered mantle** may stay in place, or it may move down hillslopes, down rivers, and down submarine slopes. This downslope movement is caused by gravity and by fluid forces. The term **mass wasting** is sometimes used to describe all processes that lower the ground surface. It is also used more specifically as a synonym of **mass movement**, which is the bulk transfer of bodies of rock debris down slopes under the influence of gravity. **Erosion**, which is derived from the Latin (*erodere*, to gnaw; *erosus*, eaten away), is the sum of all destructive processes by which weathering products are picked up (entrained) and carried by transporting media – ice, water, and wind. Most geomorphologists regard transport as an integral part of erosion, although it could be argued, somewhat pedantically, that erosion is simply the acquisition of material by mobile agencies and does not include transport. Water is a widespread transporting agent, ice far less so. Moving air may erode and carry sediments in all subaerial environments. It is most effective where vegetation cover is scanty or absent. Winds may carry sediments up slopes and over large distances (see Simonson 1995). Dust-sized particles may travel around the globe. **Denudation**, which is derived from the Latin *denudare*, meaning ‘to

Box 2.3

EARTH SYSTEM SCIENCE

Earth system science takes the view that all the terrestrial spheres interact in a basic way: the solid Earth (lithosphere, mantle, and core), atmosphere, hydrosphere, pedosphere, and biosphere are interdependent (Figure 2.3). From a geomorphological perspective, a key suggestion of this view is that denudation processes are a major link between crustal tectonic processes and the atmosphere and hydrosphere (Beaumont *et al.* 2000). Mantle convection largely drives tectonic processes, but the denudational link with the atmosphere–hydrosphere system has a large effect. In turn, tectonic processes, acting

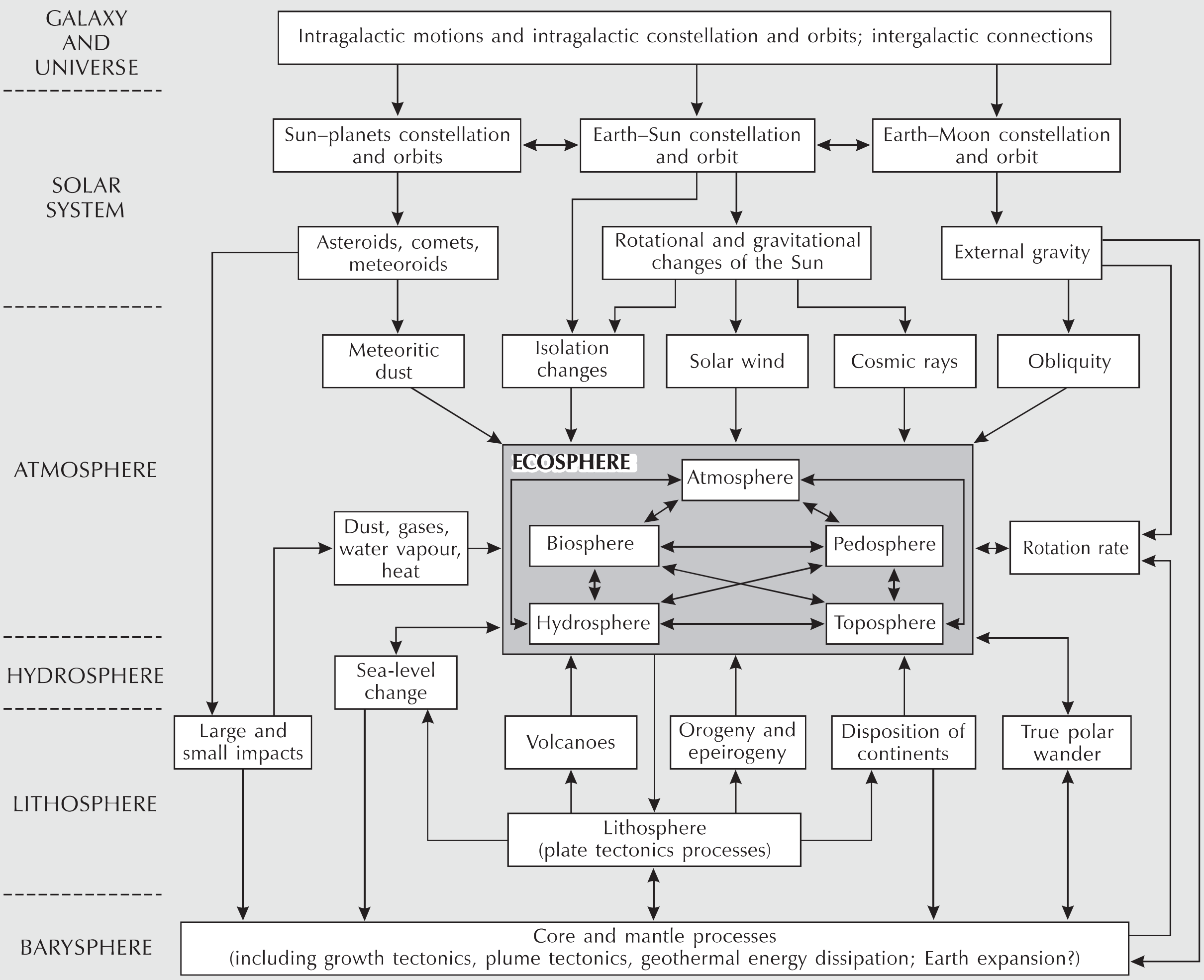


Figure 2.3 Interacting terrestrial spheres and their cosmic and geological settings.
Source: Adapted from Huggett (1991, 1995, 1997)

through the climatic effects of mountain ranges, influence the atmosphere. Similarly, the Earth's climate depends upon ocean circulation patterns, which in turn are influenced by the distribution of continents and oceans, and ultimately upon long-term changes in mantle convection.

The denudational link works through weathering, the carbon cycle, and the unloading of crustal material. Growing mountains and plateaux influence chemical weathering rates. As mountains grow, atmospheric carbon dioxide combines with the fresh rocks during weathering and is carried to the sea. Global cooling during the Cenozoic era may have been instigated by the uplift of the Tibetan plateau (p. 31). Increase in chemical weathering associated with this uplift has caused a decrease in atmospheric carbon dioxide concentrations over the last 40 million years (Raymo and Ruddiman 1992; Ruddiman 1997). The interaction of continental drift, runoff, and weathering has also affected global climates during the last 570 million years (Otto-Bliesner 1995). The removal of surface material by erosion along passive margins, as in the Western Ghats in India, causes a different effect. Unburdened by part of its surficial layers, and in conjunction with the deposition of sediment in offshore basins, the lithosphere rises by 'flexural rebound', promoting the growth of escarpments that wear back and are separated from inland plateaux that wear down (p. 110).

lay bare', is the conjoint action of weathering and erosion, which processes simultaneously wear away the land surface.

Water and ice in the pedosphere (including the weathered part of exposed rocks) may be regarded as liquid and solid components of the weathered mantle. Weathered products, along with water and ice, tend to flow downhill along lines of least resistance, which typically lie at right angles to the topographic contours. The flowlines run from mountain and hill summits to sea floors. In moving down a flowline, the relative proportion of water to sediment alters. On hillslopes, there is little, if any, water to a large body of sediment. Mass movements prevail. These take place under the influence of gravity, without the aid of moving water, ice, or air. In glaciers, rivers, and seas, a large body of water bears some suspended and dissolved sediment. Movement occurs through glacial, fluvial, and marine transport.

Deposition is the laying down of sediment by chemical, physical, or biological means. Gravitational and fluid forces move eroded material. Where the transporting capacity of the fluid is insufficient to carry the solid sediment load, or where the chemical environment leads to the precipitation of the solute load, deposition of sediment occurs. Sedimentary bodies occur where deposition outpaces erosion, and where chemical precipitation exceeds solutional loss.

THE GLOBAL PATTERN OF DENUDATION

Measurements of the amount of sediment annually carried down the Mississippi River were made during the 1840s, and Archibald Geikie worked out the rates of modern denudation in some of the world's major rivers in the 1860s. Measurements of the dissolved load of rivers enabled estimates of chemical denudation rates to be made in the first few decades of the twentieth century. Not until after the 'quantitative revolution' in geomorphology, which started in the 1940s, were rates of geomorphic processes measured in different environments and a global picture of denudation rates pieced together.

Mechanical denudation

Measuring denudation rates

Overall rates of **denudation** are judged from the dissolved and suspended loads of rivers, from reservoir sedimentation, and from the rates of geological sedimentation. Figure 2.4a depicts the pattern of sediment yield from the world's major drainage basins, and Figure 2.4b displays the annual discharge of sediment from the

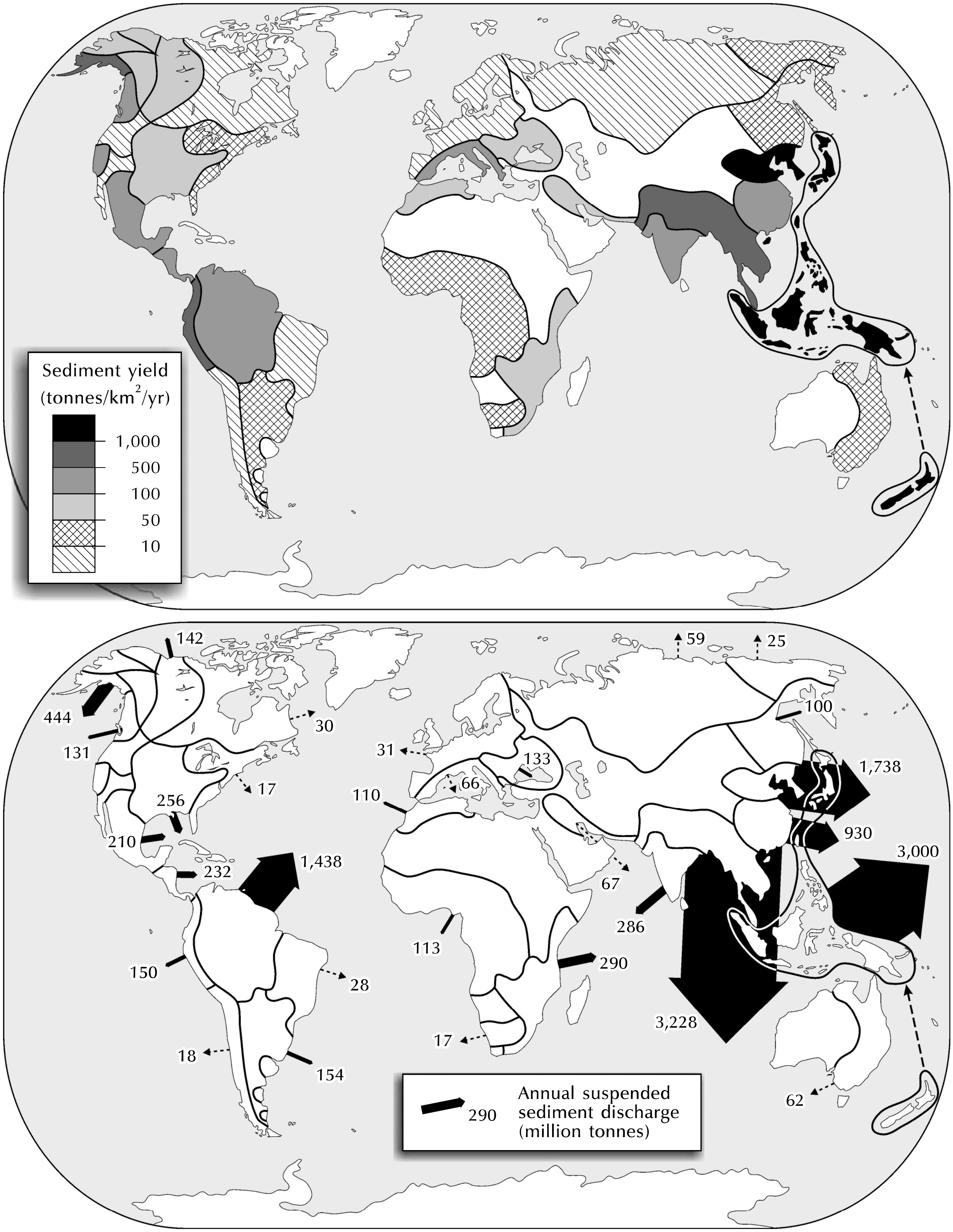


Figure 2.4 (a) Sediment yield of the world's chief drainage basins. Blank spaces indicate essentially no discharge to the ocean. (b) Annual discharge of suspended sediment from large drainage basins of the world. The width of the arrows corresponds to relative discharge. Numbers refer to average annual input in millions of tonnes. The direction of the arrows does not indicate the direction of sediment movement.
Source: Adapted from Milliman and Meade (1983)

Table 2.2 Chemical and mechanical denudation of the continents

Continent	*Chemical denudation*[a]		*Mechanical denudation*[b]		*Ratio of mechanical to chemical denudation*	*Specific discharge ($l/s/km^2$)*
	Drainage area (10^6 km^2)	*Solute yield ($t/km^2/yr$)*	*Drainage area (10^6 km^2)*	*Solute yield ($t/km^2/yr$)*		
Africa	17.55	9.12	15.34	35	3.84	6.1
North America	21.5	33.44	17.50[c]	84	2.51	8.1
South America	16.4	29.76	17.90	97	3.26	21.2
Asia	31.46	46.22	16.88	380	8.22	12.5
Europe	8.3	49.16	15.78[d]	58	1.18	9.7
Oceania	4.7	54.04	5.20	1,028[e]	19.02	16.1

Notes:
[a] Data from Meybeck (1979, annex 3)
[b] Data from Milliman and Meade (1983, Table 4)
[c] Includes Central America
[d] Milliman and Meade separate Europe (4.61×10^6 km^2) and Eurasian Arctic (11.17×10^6 km^2)
[e] The sediment yield for Australia is a mere 28 $t/km^2/yr$, whereas the yield for large Pacific islands is 1,028 $t/km^2/yr$
Source: After Huggett (1991, 87)

world's major rivers to the sea. It should be emphasized that these figures do not measure the total rate of soil erosion, since much sediment is eroded from upland areas and deposited on lowlands where it remains in store, so delaying for a long time its arrival at the sea (Milliman and Meade 1983). Table 2.2 shows the breakdown of chemical and mechanical denudation by continent.

Factors controlling denudation rates

The controls on mechanical denudation are so complex and the data so sketchy that it is challenging to attempt to assess the comparative roles of the variables involved. Undaunted, some researchers have tried to make sense of the available data (e.g. Fournier 1960; Strakhov 1967). Frédéric Fournier (1960), using sediment data from 78 drainage basins, correlated suspended **sediment yield** with a climatic parameter, p^2/P, where p is the rainfall of the month with the highest rainfall and P is the mean annual rainfall. Although, as might be expected, sediment yields increased as rainfall increased, a better degree of explanation was found when basins were grouped into relief classes. Fournier fitted an empirical equation to the data:

$$\log E = -1.56 + 2.65 \log(p^2/P + 0.46 \log \overline{H} - \tan\theta)$$

where E is suspended sediment yield ($t/km^2/yr$), p^2/P is the climatic factor (mm), $\overline{H}$ is mean height of a drainage basin, and tan θ (theta) is the tangent of the mean slope of a drainage basin. Applying this equation, Fournier mapped the distribution of world mechanical erosion. His map portrayed maximum rates in the seasonally humid tropics, declining in equatorial regions where there is no seasonal effect, and also declining in arid regions, where total runoff is low.

John D. Milliman (1980) identified several natural factors that appear to control the suspended sediment load of rivers: drainage basin relief, drainage basin area, specific discharge, drainage basin geology, climate, and the presence of lakes. The climatic factor influences suspended sediment load through mean annual temperature, total rainfall, and the seasonality of rainfall. Heavy rainfall tends to generate high runoff, but heavy seasonal rainfall, as in the monsoon climate of southern Asia, is very efficacious in producing a big load of suspended sediment. On the other hand, in areas of high, year-round rainfall, such as the Congo basin, sediment loads are not necessarily high. In arid regions, low rainfall produces little river discharge and low sediment yields; but, owing to the lack of water, suspended sediment concentrations may still be high. This is the case for many Australian rivers. The greatest suspended sediment yields come from mountainous tropical islands, areas with active glaciers, mountainous areas near coasts, and areas draining loess soils: they are not determined directly by climate (Berner and Berner 1987, 183). As one might expect, sediments deposited on inner continental shelves reflect climatic differences in source basins: mud is most abundant off areas with high temperature and high rainfall; sand is everywhere abundant but especially so in areas of moderate temperature and rainfall and in all arid areas save those with extremely cold climates; gravel is most common off areas with low temperature; and rock is most common off cold areas (Hayes 1967).

Large amounts of quartz, in association with high ratios of silica to alumina, in river sediments indicate intense tropical weathering regimes. Work carried out on the chemistry of river sediments has revealed patterns attributable to differing weathering regimes in (1) the tropical zone and (2) the temperate and frigid zones. River sands with high quartz and high silica-to-alumina ratios occur mainly in tropical river basins of low relief, where weathering is intense enough (or has proceeded uninterrupted long enough) to eliminate any differences arising from rock type, while river sands with low quartz content but high silica-to-alumina ratios occur chiefly in the basins located in temperate and frigid regions (Potter 1978). A basic distinction between tropical regions, with intense weathering regimes, and temperate and frigid regions, with less intense weathering regimes, is also brought out by the composition of the particulate load of rivers (Martin and Meybeck 1979). The tropical rivers studied had high concentrations of iron and aluminium relative to soluble elements because their particulate load was derived from soils in which soluble material had been thoroughly leached. The temperate and arctic rivers studied had lower concentrations of iron and aluminium in suspended matter relative to soluble elements because a smaller fraction of the soluble constituents had been removed. This broad pattern will almost certainly be distorted by the effects of relief and rock type. Indeed, the particulate load (p. 72) data include exceptions to the rule: some of their tropical rivers have high calcium concentrations, probably owing to the occurrence of limestone within the basin. Moreover, in explaining the generally low concentrations of calcium in sediments of tropical rivers, it should be borne in mind that carbonate rocks are more abundant in the temperate zone than in the tropical zone (cf. Figure 8.2).

Climate and denudation

Ignoring infrequent but extreme values and correcting for the effects of relief, overall rates of denudation show a relationship with climate (Table 2.3). Valley glaciation is substantially faster than normal erosion in any climate, though not necessarily so erosion by ice sheets. The wide spread of denudation rates in polar and montane environments may reflect the large range of rainfall encountered. The lowest minimum and, possibly, the lowest maximum rates of denudation occur in humid temperate climates, where creep rates are slow, wash is very slow owing to the dense cover of vegetation, and solution is relatively slow because of the low temperatures. Other conditions being the same, the rate of denudation in temperate continental climates is somewhat brisker. Semi-arid, savannah, and tropical landscapes all appear to denude fairly rapidly. Clearly, further long-term studies of denudational processes in all climatic zones are needed to obtain a clearer picture of the global pattern of denudation.

Chemical denudation

The controls on the rates of chemical denudation are perhaps easier to ascertain than the controls on the rates of mechanical denudation. Reliable estimates of the loss of material from continents in solution have been available for several decades (e.g. Livingstone 1963), though more recent estimates overcome some of the deficiencies in the older data sets. It is clear from the data in Table 2.2 that the amount of material removed in solution from continents is not directly related to the average specific discharge (discharge per unit area). South America has the highest specific discharge but the second-lowest chemical denudation rate. Europe has a relatively low specific discharge but the second-highest chemical

Table 2.3 Rates of denudation in climatic zones

Climate	*Relief*	*Typical range for denudation rate (mm/millennium)*	
		Minimum	*Maximum*
Glacial	Normal (= ice sheets)	50	200
	Steep (= valley glaciers)	1,000	5,000
Polar and montane	Mostly steep	10	1,000
Temperate maritime	Mostly normal	5	100
Temperate continental	Normal	10	100
	Steep	100	200+
Mediterranean	–	10	?
Semi-arid	Normal	100	1,000
Arid	–	10	?
Subtropical	–	10?	1,000?
Savannah	–	100	500
Tropical rainforest	Normal	10	100
	Steep	100	1,000
Any climate	Badlands	1,000	1,000,000

Source: Adapted from Saunders and Young (1983)

denudation rate. On the other hand, Africa has the lowest specific discharge and the lowest chemical denudation rate. In short, the continents show differences in resistance to being worn away that cannot be accounted for merely in terms of climatic differences.

The primary controls on chemical denudation of the continents can be elicited from data on the chemical composition of the world's major rivers (Table 2.4). The differences in solute composition of river water between continents result partly from differences of **relief** and **lithology**, and partly from **climatic differences**. Waters draining off the continents are dominated by calcium ions and bicarbonate ions. These chemical species account for the dilute waters of South America and the more concentrated waters of Europe. Dissolved silica and chlorine concentrations show no consistent relationship with total dissolved solids. The reciprocal relation between calcium ion concentrations and dissolved silica concentrations suggests a degree of control by rock type: chiefly sedimentary rocks underlie Europe and North America, whereas mainly crystalline rocks underlie Africa and South America. However, because the continents mainly consist of a heterogeneous mixture of rocks, it would be unwise to read too much into these figures and to overplay this interpretation.

Many factors affect the natural chemical composition of river water: the amount and nature of rainfall and evaporation, drainage basin geology and weathering history, average temperature, relief, and biota (Berner and Berner 1987, 193). According to Ronald J. Gibbs (1970, 1973), who plotted total dissolved solids of some major rivers against the content of calcium plus sodium, there are three chief types of surface waters:

1 Waters with low total dissolved solid loads (about 10 mg/l) but large loads of dissolved calcium and sodium, such as the Matari and Negro rivers, which depend very much on the amount and composition of precipitation.
2 Waters with intermediate total dissolved solid loads (about 100–1,000 mg/l) but low to medium loads of dissolved calcium and sodium, such as the Nile and Danube rivers, which are influenced strongly by the weathering of rocks.
3 Waters with high total dissolved solid loads (about 10,000 mg/l) and high loads of dissolved calcium

Table 2.4 Average composition of river waters by continents[a] (mg/l)

Continent	SiO_2	Ca^{2+}	Mg^{2+}	Na^+	K^+	Cl^-	SO_4^{2-}	HCO_3^-	Σi^b
Africa	12.0	5.25	2.15	3.8	1.4	3.35	3.15	26.7	45.8
North America	7.2	20.1	4.9	6.45	1.5	7.0	14.9	71.4	126.3
South America	10.3	6.3	1.4	3.3	1.0	4.1	3.5	24.4	44.0
Asia	11.0	16.6	4.3	6.6	1.55	7.6	9.7	66.2	112.5
Europe	6.8	24.2	5.2	3.15	1.05	4.65	15.1	80.1	133.5
Oceania	16.3	15.0	3.8	7.0	1.05	5.9	6.5	65.1	104.5
World	10.4	13.4	3.35	5.15	1.3	5.75	8.25	52.0	89.2

Notes:
[a] The concentrations are exoreic runoff with human inputs deducted
[b] Σi is the sum of the other materials
Source: Adapted from Meybeck (1979)

and sodium, which are determined primarily by evaporation and fractional crystallization and which are exemplified by the Rio Grande and Pecos rivers.

This classification has been the subject of much debate (see Berner and Berner 1987, 197–205), but it seems undeniable that climate does have a role in determining the composition of river water, a fact borne out by the origin of solutes entering the oceans. Chemical erosion is greatest in mountainous regions of humid temperate and tropical zones. Consequently, most of the dissolved ionic load going into the oceans originates from mountainous areas, while 74 per cent of silica comes from the tropical zone alone.

Further work has clarified the association between chemical weathering, mechanical weathering, lithology, and climate (Meybeck 1987). Chemical transport, measured as the sum of major ions plus dissolved silica, increases with increasing specific runoff, but the load for a given runoff depends on underlying rock type (Figure 2.5). Individual solutes show a similar pattern. Dissolved silica is interesting because, though the rate of increase with increasing specific discharge is roughly the same in all climates, the actual amount of dissolved silica increases with increasing temperature (Figure 2.5b). This situation suggests that, although lithology, distance to the ocean, and climate all affect solute concentration in rivers, transport rates, especially in the major rivers, depend first and foremost on specific river runoff (itself related to climatic factors) and then on lithology.

Regional and global patterns of denudation

Enormous variations in sediment and solute loads of rivers occur within particular regions owing to the local effects of rock type, vegetation cover, and so forth. Attempts to account for regional variations of denudation have met with more success than attempts to explain global patterns, largely because coverage of measuring stations is better and it is easier to take factors other than climate into consideration. Positive correlations between suspended sediment yields and mean annual rainfall and mean annual runoff have been established for drainage basins in all parts of the world, and simply demonstrate the fact that the more water that enters the system, the greater the erosivity. Solute loads, like suspended sediment loads, exhibit striking local variations about the global trend. The effects of **rock type** in particular become far more pronounced in smaller regions. For example, dissolved loads in Great Britain range from 10 to more than 200 t/km^2/yr, and the national pattern is influenced far more by lithology than by the amount of annual runoff (Walling and Webb 1986). Very high solute loads are associated with outcrops of soluble rocks. An exceedingly high solute load of 6,000 t/km^2/yr has been recorded in the River Cana, which drains an area of halite deposits in Amazonia; and a load of 750 t/km^2/yr

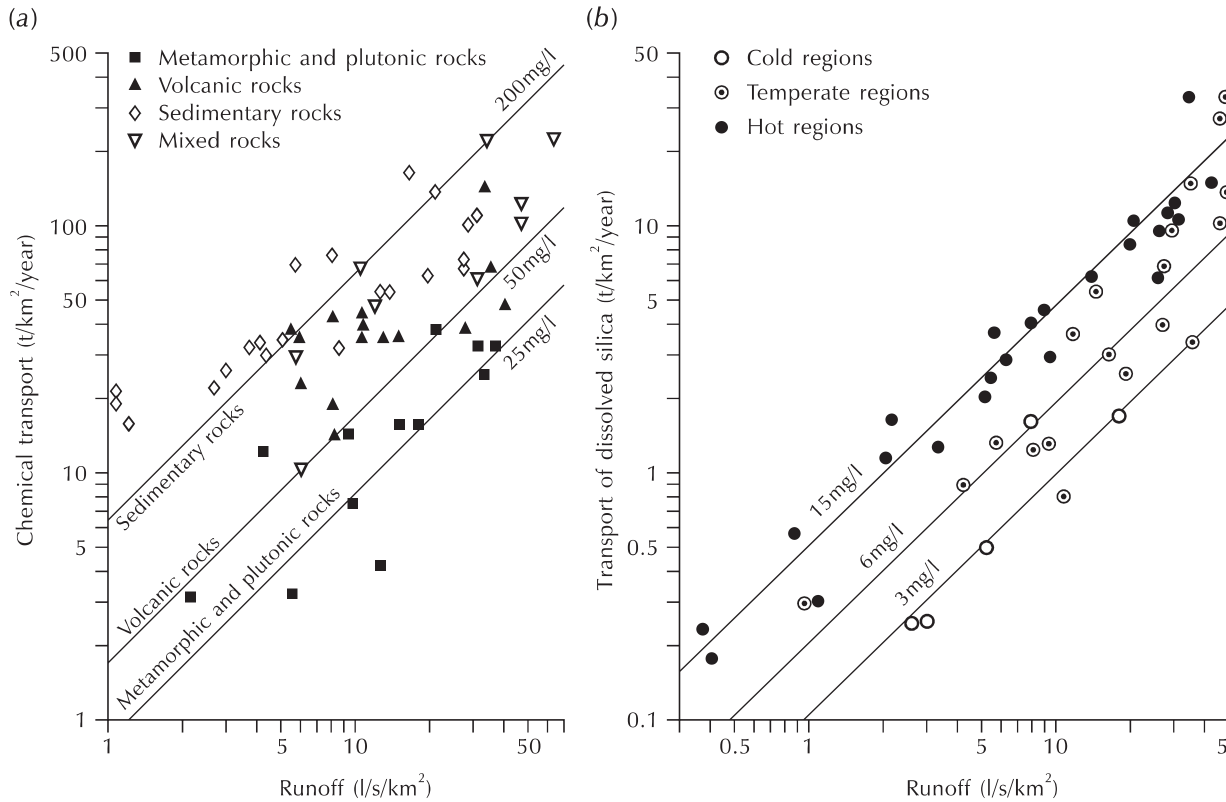


Figure 2.5 Dissolved loads in relation to runoff. (a) Chemical transport of all major ions plus dissolved silica versus runoff (specific discharge) for various major drainage basins underlain by sedimentary, volcanic, and metamorphic and plutonic rocks. (b) Evolution of the specific transport of dissolved silica for cold, temperate, and hot regions.
Source: Adapted from Meybeck (1987)

has been measured in an area draining karst terrain in Papua New Guinea.

All the general and detailed summaries of global and regional sediment yield (e.g. Fournier 1960; Jansson 1988; Milliman and Meade 1983; Summerfield and Hulton 1994) split into two camps of opinion concerning the chief determinants of erosion at large scales. Camp one sees relief as the prime factor influencing denudation rates, with climate playing a secondary role. Camp two casts climate in the leading role and relegates relief to a supporting part. Everybody seems to agree that either relief or climate, as measured by surrogates of rainfall erosivity, is the major control of erosion rates on a global scale. The problem is deciding on the relative contribution made by each factor. Jonathan D. Phillips (1990) set about the task of solving this problem by considering three questions: (1) whether indeed relief and climate are major determinants of soil loss; (2) if so, whether relief or climate is the more important determinant at the global scale; and (3) whether other factors known to influence soil loss at a local scale have a significant effect at the global scale. Phillips's results showed that slope gradient (the relief factor) is the main determinant of soil loss, explaining about 70 per cent of the maximum expected variation within global erosion rates. Climate, measured as rainfall erosivity, was less important but with relief (slope gradient) and a runoff factor accounted for 99 per cent of the maximum expected variation. The importance of a runoff factor, represented by a variable describing retention of precipitation (which is independent of climatic influences on runoff) was surprising. It was more important than the precipitation factors. Given Phillips's findings, it may pay to probe more carefully the fact that the variation in sediment yield within climatic zones is greater than the variation between climatic zones (Jansson 1988). At local scales, the influence of vegetation cover may play a critical role in dictating soil erosion rates (e.g. Thornes 1990).

Niels Hovius (1998) collated data on fourteen climatic and topographic variables used in previous studies for ninety-seven major catchments around the world. He found that none of the variables correlated well with sediment yield, which suggests that no single variable is an overriding determinant of sediment yield. However, sediment yield was successfully predicted by a combination of variables in a multiple regression equation. A five-term model explained 49 per cent of the variation in sediment yield:

$$\ln E = 3.585 - 0.416 \ln A + 4.26 \times 10^{-4} H_{max} + 0.150T + 0.095T_{range} + 0.0015R$$

where E is specific sediment yield (t/km^2/yr), A is drainage area (km^2), H_{max} is the maximum elevation of the catchment (m), T is the mean annual temperature (°C), T_{range} is the annual temperature range (°C), and R is the specific runoff (mm/yr). Of course, 51 per cent of the variation in sediment yield remains unexplained by the five-term model. One factor that might explain some of this unaccounted variation is the supply of erodible material, which, in geological terms, is largely determined by the uplift of rocks. Inputs of new matter by uplift should explain additional variation beyond that explained by the erosivity of materials.

A global survey of chemical and physical erosion data drew several interesting conclusions about the comparative roles of tectonics, the environment, and humans in explaining regional variations (Stallard 1995). Four chief points emerged from this study. First, in tectonically active mountain belts, carbonate and evaporite weathering dominates dissolved loads, and the erosion of poorly lithified sediment dominates solid loads. In such regions, human activities may increase physical erosion by orders of magnitude for short periods. About 1,000 m of uplift every million years is needed to sustain the observed chemical and physical erosion rates. Second, in old mountain belts, physical erosion is lower than in young mountain belts of comparable relief, perhaps because the weakest rocks have been stripped by earlier erosion. Third, on shields, chemical and physical erosion are very slow because weak rocks are little exposed owing to former erosion. And, finally, a basic distinction may be drawn between areas where soil development and sediment storage occur (terrains where erosion is limited by transport capacity) and areas of rapid erosion (terrains where erosion is limited by the production of fresh sediment by weathering).

THE GLOBAL TECTONIC AND CLIMATIC SYSTEMS

Since the 1990s, geomorphologists have come to realize that the **global tectonic system** and the **world climate system** interact in complex ways. The interactions give rise to fundamental changes in atmospheric circulation patterns, in precipitation, in climate, in the rate of uplift and denudation, in chemical weathering, and in sedimentation (Raymo and Ruddiman 1992; Small and Anderson 1995; Montgomery *et al.* 2001). The interaction of large-scale landforms, climate, and geomorphic processes occurs in at least three ways – through the direct effect of plate tectonic process upon topography (p. 108–15), through the direct effect of topography upon climate (and the effects of climate upon uplift), and through the indirect influence of topography upon chemical weathering rates and the concentration of atmosphere carbon dioxide.

Changes in topography, such as the uplift of mountain belts and plateaux, can influence regional climates, both by locally increasing precipitation, notably on the windward side of the barrier, and through the cooling effect of raising the ground surface to higher elevations (e.g. Ollier 2004a). Changes in topography could potentially have wide-ranging impacts if they interact with key components of the Earth's climatic system. In southern Africa, uplift of 1,000 m during the Neogene, especially in the eastern part of the subcontinent, would have reduced surface temperatures by roughly the same amount as during glacial episodes at high latitudes (Partridge 1998). The uplift of the Tibetan Plateau and its bordering mountains may have actively forced climatic change by intensifying the Asian monsoon (through altering surface atmospheric pressure owing to elevation increase), by creating a high-altitude barrier to airflow that affected the jet stream, and by encouraging inter-hemispherical exchange of heat (Liu and Ding 1998; Fang *et al.* 1999a, b). These forcings seem to have occurred around 800,000 years ago. However, oxygen isotope work on late Eocene and younger deposits in the centre of the plateau suggests that this area at least has stood at more than 4 km for about 35 million years (Rowley and Currie 2006).

Recent research shows that local and regional climatic changes caused by uplift may promote further uplift through a positive feedback loop involving the extrusion of crustal rocks (e.g. Molnar and England 1990; Hodges 2006). In the Himalaya, the Asian monsoon sheds prodigious amounts of rain on the southern flanks of the mountains. The rain erodes the rocks, which enables the fluid lower crust beneath Tibet to extrude towards the zone of erosion. Uplift results from the extrusion of rock and counterbalances the erosion, which reduces the land-surface elevation. Therefore, the extrusion process keeps the front range of the Himalaya steep, which encourages heavy monsoon rains, so completing the feedback loop (but see Ollier 2006 for a different view).

Carbon dioxide is a key factor in determining mean global temperatures. Over geological timescales (millions and tens of millions of years), atmospheric carbon dioxide levels depend upon the rate of carbon dioxide input through volcanism, especially that along mid-ocean ridges, and the rate of carbon dioxide withdrawal through the weathering of silicate rocks by carbonation, a process that consumes carbon dioxide. Given that carbon dioxide inputs through volcanism seem to have varied little throughout Earth history, it is fair to assume that variations in global chemical weathering rates should explain very long-term variations in the size of the atmospheric carbon dioxide pool. So what causes large changes in chemical weathering rates? Steep slopes seem to play a crucial role. This relatively new finding rests on the fact that weathering rates depend greatly on the amount of water passing through the weathering zone. Rates are highest on steep slopes with little or no weathered mantle and high runoff. In regions experiencing these conditions, erosional processes are more likely to remove weathered material, so exposing fresh bedrock to attack by percolating water. In regions of thick weathered mantle and shallow slopes, little water reaches the weathering front and little chemical weathering occurs. Interestingly, steep slopes characterize areas of active uplift, which also happen to be areas of high precipitation and runoff. In consequence, 'variations in rates of mountain building through geological time could affect overall rates of global chemical weathering and thereby global mean temperatures by altering the concentration of atmospheric CO_2' (Summerfield 2007, 105). If chemical weathering rates increase owing to increased tectonic uplift, then CO_2 will be drawn out of the atmosphere, but there must be some overall negative feedback in the system otherwise atmospheric CO_2 would become exhausted, or would keep on increasing and cause a runaway greenhouse effect. Neither has occurred during Earth history, and the required negative feedback probably occurs through an indirect effect of temperature on chemical weathering rates. It is likely that if global temperatures increase this will speed up the hydrological cycle and increase runoff. This will, in turn, tend to increase chemical weathering rates, which will draw down atmospheric CO_2 and thereby reduce global mean temperature. It is also possible that variations in atmospheric CO_2 concentration may directly affect chemical weathering rates, and this could provide another negative feedback mechanism.

The idea that increased weathering rates associated with tectonic uplift increases erosion and removes enough carbon dioxide from the atmosphere to control climate has its dissenters. Ollier (2004a) identified what he termed 'three misconceptions' in the relationships between erosion, weathering, and carbon dioxide. First, weathering and erosion are not necessarily concurrent processes – erosion, especially erosion in mountainous regions, may occur with little chemical alteration of rock or mineral fragments. Second, in most situations, hydrolysis and not carbonation is the chief weathering process – weathering produces clays and not carbonates. Furthermore, evidence suggests that chemical weathering rates have declined since the mid- or early Tertiary, before which time deep weathering profiles formed in broad plains. Today, deep weathering profiles form only in the humid tropics. Third, Ollier questions the accepted chronology of mountain building, which sees Tibet, the highlands of western North America, and the Andes beginning to rise about 40 million years ago, favouring instead rise over the last few million years.

SUMMARY

Three grand cycles of matter affect Earth surface processes – the water cycle (evaporation, condensation, precipitation, and runoff), the rock cycle (uplift, weathering, erosion, deposition, and lithification), and the biogeochemical cycles. Denudation encompasses weathering and erosion. Erosive agents – ice, water, and wind – pick up weathered debris, transport it, and deposit it. Climate partly determines denudation (weathering and erosion). In addition, geological and topographic factors affect mechanical denudation. Climate, rock type, topographic factors, and organisms influence chemical denudation. Climate, topography, and plate tectonic process interact in complex ways. Uplift changes climates, climatic changes may increase erosion, erosion may affect the flow of crustal rocks and so influence uplift. Erosion of mountains may affect the carbon dioxide balance of the atmosphere and promote climatic change.

ESSAY QUESTIONS

1. **To what extent are the Earth's grand 'cycles' interconnected?**
2. **How important are substrates in explaining land form?**
3. **Assess the relative importance of the factors that influence denudation rates.**

FURTHER READING

Berner, E. K. and Berner, R. A. (1987) *The Global Water Cycle: Geochemistry and Environment*. Englewood Cliffs, N.J.: Prentice Hall.
Old but still worth reading.

Berner, R. A. and Berner, E. K. (1995) *Global Environment: Water and Geochemical Cycles*. Upper Saddle River, N.J.: Prentice Hall.
A later incarnation of the previous book.

Ruddiman, W. F. (ed.) (1997) *Tectonic Uplift and Climatic Change*. New York: Plenum Press.
A detailed account of the connections between tectonics, weathering, and climate.

Westbroek, P. (1991) *Life as a Geological Force: Dynamics of the Earth*. New York: W. W. Norton.
A leisurely and winning introduction to geological and biogeochemical cycles from the perspective of Earth history.

3

GEOMORPHIC MATERIALS AND PROCESSES

Weathering, erosion, transport, and some soil processes create sediments and soils. This chapter covers:

- sediment production by weathering
- sediment transport by gravity, water, wind, the sea, and ice
- sediment deposition
- humans as geomorphic agents

Geomorphic footprint

After the Earth had evolved a solid land surface and an atmosphere, the water cycle and plate tectonic processes combined to create the rock cycle. Weathering, transport, and deposition are essential processes in this cycle. In conjunction with geological structures, tectonics processes, climate, and living things, they fashion landforms and landscapes. Over the last two centuries or so, humans have had an increasingly significant impact on the transfer of Earth materials and the modification of landforms, chiefly through agricultural practices, mining and quarrying, and the building of roads and cities. As Harrison Brown (1956, 1031) commented:

> A population of 30 billion would consume rock at a rate of about 1,500 tons per year. If we were to assume that all the land areas of the world were available for such processing, then, on the average, man [*sic*] would "eat" his way downward at a rate of 3.3 millimeters per year, or over 3 meters per millennium. This figure gives us some idea of the denudation rates that might be approached in the centuries ahead. And it gives us an idea of the powers for denudation which lie in mankind's hands.

The 'geomorphic footprint' is a measure of the rate at which humans create new landforms and mobilize sediment (Rivas *et al.* 2006). For four study areas – one in northern Spain and three in central and eastern Argentina – new landforms were created by excavation and mining activities at a rate of 7.9 m^2 per person per year in the Spanish area and 5.93 m^2 per person per year in the Argentinian areas. The volume of sediment created by these activities was 30.4 m^3 per person per year and 6.4 m^3 per person per year for the Spanish and Argentinian areas respectively. These values convert to a sediment mobilization rate of 2.4 mm/yr for the Spanish study site and 0.8 mm/yr for the Argentinian study sites, which values exceed the rate mobilization of sediment by natural processes by an order of magnitude of two (cf. Table 2.3). If these figures are typical of other human-dominated areas, then Brown's denudation rates may be reached during the present century with a smaller population.

Soils and sediments are geomorphic materials. Weathering, erosion, transport, and some soil processes create them. This chapter will outline the processes of weathering, general principles of sediment transport, the chief types of transport (fluvial, aeolian, coastal, glacial), and sediment deposition.

WEATHERING: SEDIMENT PRODUCTION

Weathering debris

Weathering acts upon rocks to produce solid, colloidal, and soluble materials. These materials differ in size and behaviour:

1 **Solids** range from boulders, through sand, and silt, to clay (Table 3.1). They are large, medium, and small fragments of rock subjected to disintegration and decomposition plus new materials, especially secondary clays built from the weathering products by a process called **neoformation**. At the lower end of the size range they grade into pre-colloids, colloids, and solutes.
2 **Solutes** are 'particles' less than 1 nm (nanometre) in diameter that are highly dispersed and exist in molecular solution.
3 **Colloids** are particles of organic and mineral substances that range in size from 1 to 100 nm. They normally exist in a highly dispersed state but may adopt a semi-solid form. Common colloids produced by weathering are oxides and hydroxides of silicon, aluminium, and iron. Amorphous silica and opaline silica are colloidal forms of silicon dioxide. Gibbsite and boehmite are aluminium hydroxides. Hematite is an iron oxide and goethite a hydrous iron oxide. **Pre-colloidal materials** are transitional to solids and range in size from about 100 to 1,000 nm.

Mechanical or physical weathering

Mechanical processes reduce rocks into progressively smaller fragments. The disintegration increases the surface area exposed to chemical attack. The main processes of **mechanical weathering** are unloading, frost action, thermal stress caused by heating and cooling, swelling and shrinking due to wetting and drying, and pressures exerted by salt-crystal growth. A significant ingredient in mechanical weathering is **fatigue**, which is the repeated generation of stress, by for instance heating and cooling, in a rock. The result of fatigue is that the rock will fracture at a lower stress level than a non-fatigued specimen.

Unloading

When erosion removes surface material, the confining pressure on the underlying rocks is eased. The lower pressure enables mineral grains to move further apart, creating voids, and the rock expands or dilates. In mineshafts cut in granite or other dense rocks, the pressure release can cause treacherous explosive **rockbursts**. Under natural conditions, rock dilates at right-angles to an erosional surface (valley side, rock face, or whatever). The dilation produces large or small cracks (fractures and joints) that run parallel to the surface. The dilation joints encourage rock falls and other kinds of mass movement. The small fractures and incipient joints provide lines of weakness along which individual crystals or particles may disintegrate and exfoliation may occur. **Exfoliation** is the spalling of rock sheets from the main rock body. In some

Table 3.1 Size grades of sedimentary particles

Particle names		*Particle diameter*		*Deposits*	
		ϕ (phi) units[a]	*mm*	*Unconsolidated examples*	*Consolidated examples*
Gravel[b]	Boulders	<-8	>256	*Rudaceous deposits*	
	Cobbles	-6 to -8	64–256	Till	Conglomerate, breccia, gritstone
	Pebbles	-2 to -6	4–64		
	Granules	-1 to -2	2–4		
Sand	Very coarse sand	0 to -1	1–2	*Arenaceous deposits*	
	Coarse sand	1 to 0	0.5–1	Sand	Sandstone, arkose, greywacke, flags
	Medium sand	2 to 1	0.25–0.5		
	Fine sand	3 to 2	0.125–0.25		
	Very fine sand	4 to 3	0.0625–0.125		
Silt		8 to 4	0.002–0.0625	*Argillaceous deposits*	
Clay		>8	<0.002	Clay, mud, silt	Siltstone, claystone, mudstone, shale, marl

Notes:
[a] The phi scale expresses the particle diameter, *d*, as the negative logarithm to the base 2: $\phi = -\log_2 d$
[b] The subdivisions of coarse particles vary according to authorities

rocks, such as granite, it may produce convex hills known as **exfoliation domes**.

Frost action

Water occupying the pores and interstices within a soil or rock body expands upon freezing by 9 per cent. This expansion builds up stress in the pores and fissures, causing the physical disintegration of rocks. **Frost weathering** or **frost shattering** breaks off small grains and large boulders, the boulders then being fragmented into smaller pieces. It is an important process in cold environments, where **freeze–thaw cycles** are common. Furthermore, if water-filled fissures and pores freeze rapidly at the surface, the expanding ice induces a hydrostatic or cryostatic pressure that is transmitted with equal intensity through all the interconnected hollow spaces to the still unfrozen water below. The force produced is large enough to shatter rocks, and the process is called **hydrofracturing** (Selby 1982, 16). It means that frost shattering can occur below the depth of frozen ground. In unsaturated soils, once the water is frozen, the water vapour circulating through the still open pores and fissures that comes into contact with the ice condenses and freezes. The result is that ice lenses grow that push up the overlying layers of soil. This process is called **frost heaving** and is common in glacial and periglacial environments (cf. p. 66).

Heating and cooling

Rocks have low thermal conductivities, which means that they are not good at conducting heat away from their surfaces. When they are heated, the outer few millimetres become much hotter than the inner portion and the outsides expand more than the insides. In addition, in rocks composed of crystals of different colours, the darker crystals warm up faster and cool down more slowly than the lighter crystals. All these thermal stresses may cause rock disintegration and the formation of rock flakes, shells, and huge sheets. Repeated heating and cooling produces a fatigue effect, which enhances the **thermal weathering**. The production of sheets by thermal stress was once called exfoliation, but today exfoliation encompasses a wider range of processes that produce rock flakes and rock sheets of various kinds and sizes. Intense heat generated by bush fires and nuclear explosions assuredly may cause rock to flake and split. In India and Egypt, fire was for many years used as a quarrying tool. However, the everyday temperature fluctuations found even in deserts are well below the extremes achieved by local fires. Recent research points to chemical, not physical, weathering as the key to understanding rock disintegration, flaking, and splitting. In the Egyptian desert near Cairo, for instance, where rainfall is very low and temperatures very high, fallen granite columns are more weathered on their shady sides than they are on the sides exposed to the Sun (Twidale and Campbell 1993, 95). Also, rock disintegration and flaking occur at depths where daily heat stresses would be negligible. Current opinion thus favours moisture, which is present even in hot deserts, as the chief agent of rock decay and rock breakdown, under both humid and arid conditions.

Wetting and drying

Some **clay minerals** (Box 3.1), including smectite and vermiculite, swell upon wetting and shrink when they dry out. Materials containing these clays, such as mudstone and shale, expand considerably on wetting, inducing microcrack formation, the widening of existing cracks, or the disintegration of the rock mass. Upon drying, the absorbed water of the expanded clays evaporates, and shrinkage cracks form. Alternate swelling and shrinking associated with wetting–drying cycles, in conjunction with the fatigue effect, leads to **wet–dry weathering**, or **slaking**, which physically disintegrates rocks.

Salt-crystal growth

In coastal and arid regions, crystals may grow in saline solutions on evaporation. Salt crystallizing within the interstices of rocks produces stresses, which widen them, and this leads to granular disintegration. This process

Box 3.1

CLAY MINERALS

Clay minerals are hydrous silicates that contain metal cations. They are variously known as **layer silicates**, **phyllosilicates**, and **sheet silicates**. Their basic building blocks are sheets of silica (Si) **tetrahedra** and oxygen (O) and hydroxyl (OH) **octahedra**. A silica tetrahedron consists of four oxygen atoms surrounding a silicon atom. Aluminium frequently, and iron less frequently, substitutes for the silicon. The tetrahedra link by sharing three corners to form a hexagon mesh pattern. An oxygen–hydroxyl octahedron consists of a combination of hydroxyl and oxygen atoms surrounding an aluminium (Al) atom. The octahedra are linked by sharing edges. The silica sheets and the octahedral sheets share atoms of oxygen, the oxygen on the fourth corner of the tetrahedrons forming part of the adjacent octahedral sheet.

Three groups of clay minerals are formed by combining the two types of sheet (Figure 3.1). The **1 : 1 clays** have one tetrahedral sheet combined with one flanking octahedral sheet, closely bonded by hydrogen ions (Figure 3.1a). The anions exposed at the surface of the octahedral sheets are hydroxyls. **Kaolinite** is an example, the structural formula of which is $Al_2Si_2O_5(OH)_4$. Halloysite is similar in composition to kaolinite. The **2 : 1 clays** have an octahedral sheet with two flanking tetrahedral sheets, which are strongly bonded by potassium ions (Figure 3.1b). An example is **illite**. A third group, the **2 : 2 clays**, consist of 2 : 1 layers with octahedral sheets between them (Figure 3.1c). An example is **smectite** (formerly called **montmorillonite**), which is similar to illite but the layers are deeper and allow water and certain organic substances to enter the lattice leading to expansion or swelling. This allows much ion exchange within the clays.

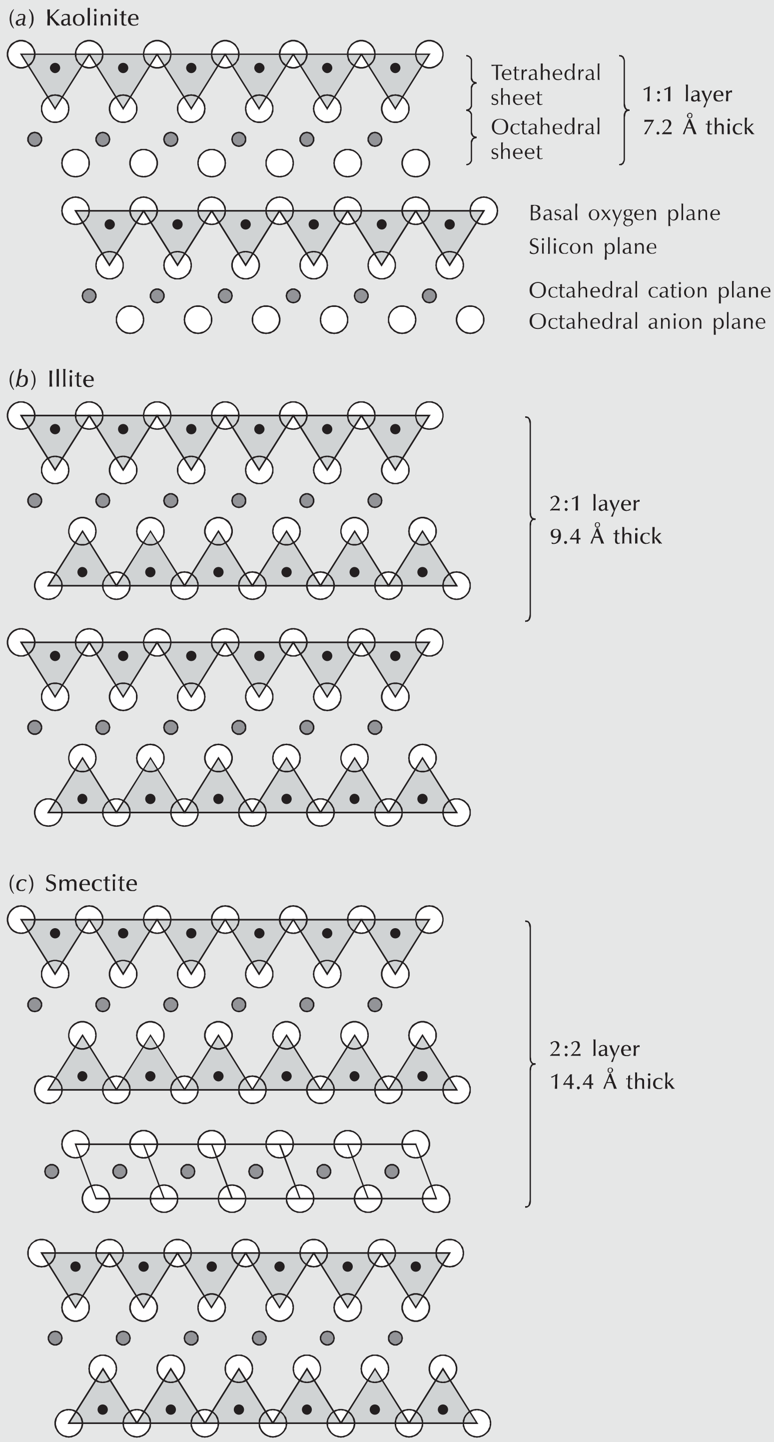


Figure 3.1 Clay mineral structure. (a) Kaolinite, a 1 : 1 dioctahedral layer silicate. (b) Illite, a 2 : 1 layer silicate, consisting of one octahedral sheet with two flanking tetrahedral sheets. (c) Smectite, a 2 : 2 layer silicate, consisting of 2 : 1 layers with octahedral sheets between. Å stands for an angstrom, a unit of length ($1Å = 10^{-8}$ cm).
Source: After Taylor and Eggleton (2001, 59, 61)

is known as **salt weathering** (Wellman and Wilson 1965). When salt crystals formed within pores are heated, or saturated with water, they expand and exert pressure against the confining pore walls; this produces thermal stress or hydration stress respectively, both of which contribute to salt weathering.

Chemical weathering

Weathering involves a huge number of chemical reactions acting together upon many different types of rock under the full gamut of climatic conditions. Six main chemical reactions are engaged in rock decomposition: solution, hydration, oxidation and reduction, carbonation, and hydrolysis.

Solution

Mineral salts may dissolve in water, which is a very effective solvent. The process, which is called **solution** or **dissolution**, involves the dissociation of the molecules into their anions and cations and each ion becomes surrounded by water. It is a mechanical rather than a chemical process, but is normally discussed with chemical weathering as it occurs in partnership with other chemical weathering processes. Solution is readily reversed – when the solution becomes saturated some of the dissolved material precipitates. The saturation level is defined by the equilibrium solubility, that is, the amount of a substance that can dissolve in water. It is expressed as parts per million (ppm) by volume or milligrams per litre (mg/l). Once a solution is saturated, no more of the substance can dissolve. Minerals vary in their solubility. The most soluble natural minerals are chlorides of the alkali metals: rock salt or halite (NaCl) and potash salt (KCl). These are found only in very arid climates. Gypsum ($CaSO_4.2H_2O$) is also fairly soluble, as is limestone. Quartz has a very low solubility. The solubility of many minerals depends upon the number of free hydrogen ions in the water, which may be measured as the pH value (Box 3.2).

Hydration

Hydration is transitional between chemical and mechanical weathering. It occurs when minerals absorb water molecules on their edges and surfaces, or, for simple salts, in their crystal lattices, without otherwise changing the chemical composition of the original material. For instance, if water is added to anhydrite, which is calcium sulphate ($CaSO_4$), gypsum ($CaSO_4.2H_2O$) is produced. The water in the crystal lattice leads to an increase of volume, which may cause hydration folding in gypsum sandwiched between other beds. Under humid mid-latitude climates, brownish to yellowish soil colours are caused by the hydration of the reddish iron oxide hematite to rust-coloured goethite. The taking up of water by clay particles is also a form of hydration. It leads to the clay's swelling when wet. Hydration assists other weathering processes by placing water molecules deep inside crystal structures.

Oxidation and reduction

Oxidation occurs when an atom or an ion loses an electron, increasing its positive charge or decreasing its negative charge. It involves oxygen combining with a substance. Oxygen dissolved in water is a prevalent oxidizing agent in the environment. **Oxidation weathering** chiefly affects minerals containing iron, though such elements as manganese, sulphur, and titanium may also be oxidized. The reaction for iron, which occurs mainly when oxygen dissolved in water comes into contact with iron-containing minerals, is written:

$$4Fe^{2+} + 3O_2 + 2e \rightarrow 2Fe_2O_3 \; [e = \text{electron}]$$

Alternatively, the ferrous iron, Fe^{2+}, which occurs in most rock-forming minerals, may be converted to its ferric form, Fe^{3+}, upsetting the neutral charge of the crystal lattice, sometimes causing it to collapse and making the mineral more prone to chemical attack.

If soil or rock becomes saturated with stagnant water, it becomes oxygen-deficient and, with the aid of **anaerobic bacteria**, reduction occurs. **Reduction** is the opposite of oxidation, and the changes it promotes are called gleying. In colour, gley soil horizons are commonly a shade of grey.

The propensity for oxidation or reduction to occur is shown by the redox potential, Eh. This is measured in units of millivolts (mV), positive values registering

Box 3.2

pH AND Eh

pH is a measure of the **acidity** or **alkalinity** of aqueous solutions. The term stands for the concentration of hydrogen ions in a solution, with the p standing for *Potenz* (the German word for 'power'). It is expressed as a logarithmic scale of numbers ranging from about 0 to 14 (Figure 3.2). Formulaically, $pH = -\log[H^+]$, where $[H^+]$ is the hydrogen ion concentration (in gram-equivalents per litre) in an aqueous solution. A pH of 14 corresponds to a hydrogen ion concentration of 10^{-14} gram-equivalents per litre. A pH of 7, which is neutral (neither acid nor alkaline), corresponds to a hydrogen ion concentration of 10^{-7} gram-equivalents per litre. A pH of 0 corresponds to a hydrogen ion concentration of 10^{-0} (= 1) gram-equivalents per litre. A solution with a pH greater than 7 is said to be alkaline, whereas a solution with a pH less than 7 is said to be acidic (Figure 3.2). In weathering, any precipitation with a pH below 5.6 is deemed to be acidic and referred to as '**acid rain**'.

The solubility of minerals also depends upon the **Eh** or **redox** (**reduction–oxidation**) **potential** of a solution. The redox potential measures the oxidizing or reducing characteristics of a solution. More specifically, it measures the ability of a solution to supply electrons to an oxidizing agent, or to take up electrons from a reducing agent. So redox potentials are electrical potentials or voltages. Solutions may have positive or negative redox potentials, with values ranging from about −0.6 volts to +1.4 volts. High Eh values correspond to oxidizing conditions, while low Eh values correspond to reducing conditions.

Combined, pH and Eh determine the solubility of clay minerals and other weathering products. For example, goethite, a hydrous iron oxide, forms where Eh is relatively high and pH is medium. Under high oxidizing conditions (Eh > +100 millivolts) and a moderate pH, it slowly changes to hematite.

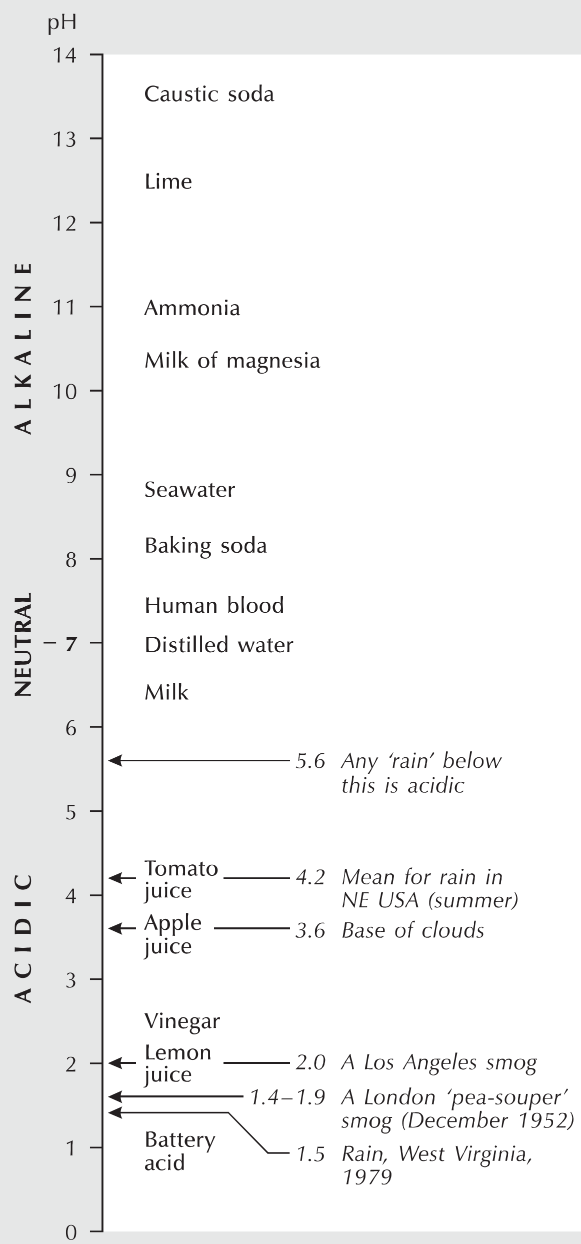


Figure 3.2 The pH scale, with the pH of assorted substances shown.

as oxidizing potential and negative values as reducing potential (Box 3.2).

Carbonation

Carbonation is the formation of carbonates, which are the salts of carbonic acid (H_2CO_3). Carbon dioxide dissolves in natural waters to form carbonic acid. The reversible reaction combines water with carbon dioxide to form carbonic acid, which then dissociates into a hydrogen ion and a bicarbonate ion. Carbonic acid attacks minerals, forming carbonates. Carbonation dominates the weathering of calcareous rocks (limestones and dolomites) where the main mineral is calcite or calcium carbonate ($CaCO_3$). Calcite reacts with carbonic acid to form calcium hydrogen carbonate ($Ca(HCO_3)_2$) that, unlike calcite, is readily dissolved in water. This is why some limestones are so prone to solution (p. 188). The reversible reactions between carbon dioxide, water, and calcium carbonate are complex. In essence, the process may be written:

$$CaCO_3 + H_2O + CO_2 \Leftrightarrow Ca^{2+} + 2HCO_3^-$$

This formula summarizes a sequence of events starting with dissolved carbon dioxide (from the air) reacting speedily with water to produce carbonic acid, which is always in an ionic state:

$$CO_2 + H_2O \Leftrightarrow H^+ + HCO_3$$

Carbonate ions from the dissolved limestone react at once with the hydrogen ions to produce bicarbonate ions:

$$CO_3{}^{2-} + H^+ \Leftrightarrow HCO_3{}^{2-}$$

This reaction upsets the chemical equilibrium in the system, more limestone goes into solution to compensate, and more dissolved carbon dioxide reacts with the water to make more carbonic acid. The process raises the concentration by about 8 mg/l, but it also brings the carbon dioxide partial pressure of the air (a measure of the amount of carbon dioxide in a unit volume of air) and in the water into disequilibrium. In response, carbon dioxide diffuses from the air to the water, which enables further solution of limestone through the chain of reactions. Diffusion of carbon dioxide through water is a slow process compared with the earlier reactions and sets the limit for limestone solution rates. Interestingly, the rate of reaction between carbonic acid and calcite increases with temperature, but the equilibrium solubility of carbon dioxide decreases with temperature. For this reason, high concentrations of carbonic acid may occur in cold regions, even though carbon dioxide is produced at a slow rate by organisms in such environments.

Carbonation is a step in the complex weathering of many other minerals, such as in the hydrolysis of feldspar.

Hydrolysis

Generally, **hydrolysis** is the main process of chemical weathering and can completely decompose or drastically modify susceptible primary minerals in rocks. In hydrolysis, water splits into **hydrogen cations** (**H^+**) and **hydroxyl anions** (**OH^-**) and reacts directly with silicate minerals in rocks and soils. The hydrogen ion is exchanged with a metal cation of the silicate minerals, commonly potassium (K^+), sodium (Na^+), calcium (Ca^{2+}), or magnesium (Mg^{2+}). The released cation then combines with the hydroxyl anion. The reaction for the hydrolysis of orthoclase, which has the chemical formula $KAlSi_3O_8$, is as follows:

$$2KAlSi_3O_8 + 2H^+ + 2OH^- \rightarrow 2HAlSi_3O_8 + 2KOH$$

So the orthoclase is converted to aluminosilicic acid, $HAlSi_3O_8$, and potassium hydroxide, KOH. The aluminosilicic acid and potassium hydroxide are unstable and react further. The potassium hydroxide is carbonated to potassium carbonate, K_2CO_3, and water, H_2O:

$$2KOH + H_2CO_3 \rightarrow K_2CO_3 + 2H_2O$$

The potassium carbonate so formed is soluble in and removed by water. The aluminosilicic acid reacts with water to produce kaolinite, $Al_2Si_2O_5(OH)_4$

(a clay mineral), and silicic acid, H_4SiO_4:

$$2HAlSi_3O_8 + 9H_2O \rightarrow Al_2Si_2O_5(OH)_4 + 2H_4SiO_4$$

The silicic acid is soluble in and removed by water leaving kaolinite as a residue, a process termed **desilication** as it involves the loss of silicon. If the solution equilibrium of the silicic acid changes, then silicon dioxide (silica) may be precipitated out of the solution:

$$H_4SiO_4 \rightarrow 2H_2O + SiO_2$$

Weathering of rock by hydrolysis may be complete or partial (Pedro 1979). **Complete hydrolysis** or **allitization** produces gibbsite. **Partial hydrolysis** produces either 1 : 1 clays by a process called **monosiallitization**, or 2 : 1 and 2 : 2 clays through a process called **bisiallitization** (cf. pp. 159–60).

Chelation

This is the removal of metal ions, and in particular ions of aluminium, iron, and manganese, from solids by binding with such organic acids as fulvic and humic acid to form soluble **organic matter–metal complexes**. The chelating agents are in part the decomposition products of plants and in part secretions from plant roots. Chelation encourages chemical weathering and the transfer of metals in the soil or rock.

Biological weathering

Some organisms attack rocks mechanically, or chemically, or by a combination of mechanical and chemical processes.

Plant roots, and especially tree roots, growing in bedding planes and joints have a **biomechanical effect** – as they grow, mounting pressure may lead to rock fracture. Dead lichen leaves a dark stain on rock surfaces. The dark spots absorb more thermal radiation than the surrounding lighter areas, so encouraging **thermal weathering**. A pale crust of excrement often found below birds' nests on rock walls reflects solar radiation and reduces local heating, so reducing the strength of rocks. In coastal environments, marine organisms bore into rocks and graze them (e.g. Yatsu 1988, 285–397; Spencer 1988; Trenhaile 1987, 64–82). This process is particularly effective in tropical limestones. Boring organisms include bivalve molluscs and clinoid sponges. An example is the blue mussel (*Mytilus edulis*). Grazing organisms include echinoids, chitons, and gastropods, all of which displace material from the rock surface. An example is the West Indian top shell (*Cittarium pica*), a herbivorous gastropod.

Under some conditions, bacteria, algae, fungi, and lichens may chemically alter minerals in rocks. The boring sponge (*Cliona celata*) secretes minute amounts of acid to bore into calcareous rocks. The rock minerals may be removed, leading to **biological rock erosion**. In an arid area of southern Tunisia, weathering is concentrated in topographic lows (pits and pans) where moisture is concentrated and algae bore, pluck, and etch the limestone substrate (Smith *et al.* 2000).

Humans have exposed bedrock in quarries, mines, and road and rail cuts. They have disrupted soils by detonating explosive devices, and they have sealed the soil in urban areas under a layer of concrete and tarmac. Their agriculture practices have greatly modified soil and weathering processes in many regions.

SEDIMENT TRANSPORT

A river in flood demonstrates sediment transport, the dirty floodwaters bearing a burden of material derived from the land surface. As well as the visible sediment, the river also carries a load of material in solution. Geomorphologists often distinguish between sediment transport, which is essentially mechanical, and solutional transport, which is essentially chemical; they also discriminate between processes involving a lot of sediment moving *en masse* – mass movement – and sediment moving as individual grains more or less dispersed in a fluid – fluid transport (cf. Statham 1977, 1). In mass movement, the weight of sediment is a key controlling factor of motion, whereas in fluid transport the action of an external fluid agency (wind or water) is the key factor. However, the distinction blurs in case of slow mass movements, which resemble flows, and in the continuous transition from dry moving material to muddy water.

Transport mechanics

Geomorphic forces

The transport of all materials, from solid particles to dissolved ions, needs a force to start and maintain motion. Such forces make boulders fall from cliffs, soils and sediment move hillslopes, and water and ice flow along channels. For this reason, the mechanical principles controlling movement underpin the understanding of transport processes (Box 3.3).

The forces that drive sediment movement largely derive from gravity, from climatic effects (heating and cooling, freezing and thawing, winds), and from the action of animals and plants. They may act directly, as in the case of gravity, or indirectly through such agencies as water and wind. In the first case, the force makes the sediment move, as in landslides; while, in the second case, the force makes the agency move (water for instance) and in turn the moving agency exerts a force on the sediment and tends to move it, as in sediment transport in rivers. The chief forces that act upon geomorphic materials are gravitational forces, fluid forces, water pressure forces, expansion forces, global fluid movements, and biological forces.

1 **Gravitational forces**. Gravity is the largest force for driving geomorphic processes. It acts directly on bodies of rock, sediment, water, and ice, tending to make them move. Moreover, it acts the world over at a nearly uniform magnitude of 9.81 metres per second per second (m/s^2), with slight variations resulting from distance from the Earth's centre and latitude.
2 **Fluid forces**. Water flows over sloping land surfaces. It does so as a subdivided or uniform sheet or as channel flows in streams and rivers. Water is a fluid so that it moves in the direction of any force that is applied to it, and no critical force is necessary. So water flows downhill under the influence of its

Box 3.3

FORCE AND RESISTANCE

A body will not move unless a force is applied, and its movement will not continue without the sustained exertion of a force. Likewise, forces act on a body at rest that are in balance while the body remains stationary. For this reason, forces are immensely important in determining if the transport of sediments takes place.

A **force** is an action in a specified direction that tends to alter the state of motion of a body. An equal and opposite force called the reaction always balances it. A boulder resting on the ground exerts a vertical force on the ground due to its weight; the ground exerts a force of the same magnitude in the opposite direction on the boulder; and, if it did not do so, the boulder would sink into the ground. Forces result from the acceleration of a body. If a body is not subject to an acceleration, then it cannot exert a force in any direction. At the Earth's surface, most bodies are subject to the acceleration due to gravity and exert a force in the direction of gravity, which is approximately vertically. The magnitude of this force is generally the weight of the body in a static condition (but, if the body is moving, the force alters).

Forces have direction and magnitude. If two or more forces are acting on a body, then the magnitude and direction of a resultant force is determinable. For example, a sediment grain entrained in flowing water is subject to several forces: a vertical force pushing it vertically upwards in the flow, the force of its own weight dragging it down vertically, and the downstream force of the flowing water carrying it along the river channel. The magnitude and direction of all these forces dictate the net direction in which the grain will travel and so whether it will stay suspended or sink to the riverbed. If a single force is known, its effects in different directions (its components) can be worked out. Take the case of a boulder on a hillslope (Figure 3.3). The weight of the boulder acts vertically in the direction of gravity, but the reaction with the ground surface prevents the boulder from moving in that direction. Nonetheless, movement downslope is possible because the weight of the boulder is resolvable into two forces – a force normal to the slope, which tends to hold the boulder in place, and a force parallel to the slope, which tends to move the boulder downhill. Normal and parallel reaction forces balance these. Now, the boulder will not move unless the downslope force can overcome the resistance to movement (friction) to counter the parallel reaction force. Once the downslope force exceeds the surface resistance, the boulder will accelerate, and its reaction then involves an inertia force due to the boulder's accelerating down the slope. This means that a smaller downslope force component is required to continue the motion at constant velocity, in the same way that it is easier to pull a sledge once it is moving than it is to start it moving.

Resistance is fundamental to transport processes. Without resistance, Earth surface materials would move under the force of gravity until the landscape was all but flat. Many factors affect resistance, but none so much as friction. Friction exists between bodies and the surface over which they move. It occurs between where matter in any state (solid, liquid, gas) come into contact, as in solids on solids, solids on fluids, fluids on fluids, and gases on solids or fluids. In a river, friction occurs at the fluid bed contact and within

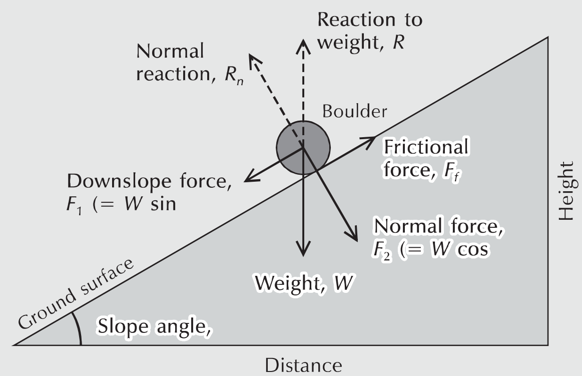


Figure 3.3 Forces acting upon a boulder lying on a hillside.

the water, owing to differential velocity of flow and turbulent eddies. In the case of a boulder at rest on a flat surface, if no lateral force is applied to the boulder, then the frictional resistance is zero as there is no force to resist. If a lateral force, F, is applied, then the frictional force, F_f, increases to balance the force system. At a critical value for F, the frictional resistance, generated between the boulder and the surface, will be unable to balance the applied force and the boulder will start to accelerate. For any given surface contact

$$F_{\text{critical}}/R = \text{a constant} = \mu_s$$

As the ratio is constant, the force required to move the boulder increases in proportion with R (the normal reaction, which, on a flat surface, is equal to the weight of the boulder).

own weight, which is a gravitational force. Moving the water uses only part of the downslope force, and the portion left after overcoming various resistances to flow may carry material in the flow or along the water–ground contact. The water also carries dissolved material that travels at the same velocity as the water and essentially behaves as part of the fluid itself.

3 **Water pressure forces**. Water in soil and sediment creates various forces that can affect sediment movement. The forces in saturated (all the pores filled) and unsaturated (some of the pores filled) differ. First, under saturated conditions with the soil or sediment immersed in a body of water (for example, below the water table), an upward buoyancy or **water pressure force** equal to the

weight of water displaces and relieves some of the downward force created by the weight of the sediment. Second, under unsaturated conditions, a **negative pore pressure** or **suction force** tends to hold the water within the pores and even draw it up from the water table by capillary rise. Such negative pore pressure increases the normal force between sediment grains and increases their resistance to movement. This capillary cohesion force keeps sandcastles from collapsing. Falling raindrops also create a force when they strike the ground. Depending on their size and terminal velocity, they may create a force strong enough to move sediment grains.

4 **Expansion forces**. Sediments, soils, and even solid rock may expand and contract in response to changes of temperature (heating and cooling, freezing and thawing) or moisture content (wetting and drying), and sometimes in response to chemical changes in minerals. Expansion tends to act equally in all directions, and so any movement that occurs is reversible. However, on slopes, the action of gravity means that expansion in a downslope direction greater than contraction in an upslope direction produces an overall downslope movement of material.

5 **Global fluid movements**. Wind carries water sediment in much the same way as water does – along the 'bed' or in suspension. But, as air is far less dense a fluid than water, for the same flow velocity it carries sediment of smaller grain size.

6 **Biological forces**. Animals and plants create forces that influence sediment movement. Plant root systems push material aside, and if this occurs on a slope an overall downslope movement may result. Burrowing animals mine soils and sediment, redistributing it across the land surface (see Butler 1995). Where animals burrow into slopes, a tendency for an overall downslope movement occurs.

In summary, most movements of sediment require a downslope force resulting from action of gravity, but climatic, meteorological, and biotic factors may also play an important role in moving materials.

Shear stress, friction, cohesion, and shear strength

A handful of key mechanisms explain much about transport processes – force, stress, friction, and shear strength. The case of soil resting on a slope demonstrates these mechanisms. The force of gravity acts upon the sediment, creating stresses. The normal stress (acting perpendicular to the slope) tends to hold the sediment in place. The **shear stress** acts in a downslope direction and, if large enough, will move the soil downhill.

Three factors resist this downhill movement – friction, cohesion, and shear strength. **Friction** resists sliding. Many factors affect it, the most important being:

- friction between the sediment and the underlying rock
- internal friction of grains within the sediment (which depends upon their size, shape, arrangement, resistance to crushing, and the number of contacts per unit volume)
- normal stress (the larger this is, the greater the degree of friction)
- smoothness of the plane of contact between the sediment and the rock, which influences the angle of friction.

A soil mass on a slope needs no externally applied force for it to move. If the slope angle is steep enough, the downslope component of the soil's weight will provide sufficient downslope force to cause movement. When the slope angle reaches a critical value, the soil will start to slide. This critical angle is the static angle of sliding friction, φ_μ, the tangent of which is equal to the coefficient of static friction. The effective normal stress, which allows for the pore water pressure in the soil, also influences sliding. In dry material, the effective normal stress is the same as the normal stress, but in wet but unsaturated soils, where pore water pressure is negative, the effective shear stress is less than the shear stress. **Cohesion** of the soil (the degree to which the individual grains are held together) also affects sliding, cohesive sediment resisting sliding more than non-cohesive sediment. Finally, **shear strength**, which is the resistance of the soil to shear stress, affects movement. Mohr–Coulomb's law relates shear strength to cohesion, gravity, and friction (see below). When shear stress (a driving force) exceeds shear strength (a resisting force), then slope failure occurs and the soil moves. In rock, weathering (which may increase cohesion), the presence of joints and bedding planes (which may reduce the angle of friction), pore water (which reduces effective normal stress and increases cohesion), and vegetation (which increases the angle of friction and may increase cohesion) affect shear strength. Other factors influencing shear strength include extra weight added to a slope as water or building materials, earthquakes, and erosion or excavation of rock units.

GRAVITATIONAL PROCESSES

Stress and strain in soils and sediments

Earth materials are subject to stress and strain. A **stress** is any force that tends to move materials downslope. Gravity is the main force, but swelling and shrinking, expansion and contraction, ice-crystal growth, and the activities of animals and plants also set up forces in a soil body. The stress of a body of soil on a slope depends largely upon the mass of the soil body, m, and the angle of slope, θ (theta):

$$\text{Stress} = m \sin \theta$$

Strain is the effect of stress upon a soil body. It may be spread uniformly throughout the body, or it may focus around joints where fracture may occur. It may affect individual particles or the entire soil column.

Materials possess an inherent resistance against downslope movement. Friction is a force that acts against gravity and resists movement. It depends on the roughness of the plane between the soil and the underlying material. Downslope movement of a soil body can occur only when the applied stress is large enough to overcome the maximum frictional resistance. Friction is expressed as a coefficient, μ (mu), which is equal to the angle at which sliding begins (called the **angle of plane sliding friction**). In addition to friction, cohesion between particles resists downslope movement. Cohesion measures the tendency of particles within the soil body to stick together. It arises through capillary suction of water in pores, compaction (which may cause small grains to interlock), chemical bonds (mainly **Van der Waals bonds**), plant root systems, and the presence of such cements as carbonates, silica, and iron oxides. Soil particles affect the mass cohesion of a soil body by tending to stick together and by generating friction between one another, which is called the **internal friction** or **shearing resistance** and is determined by particle size and shape, and the degree to which particles touch each other. The **Mohr–Coulomb** equation defines the shear stress that a body of soil on a slope can withstand before it moves:

$$\tau_s = c + \sigma \tan \phi$$

where τ_s (tau-s) is the **shear strength** of the soil, c is soil cohesion, σ (sigma) is the **normal stress** (at right-angles to the slope), and ϕ (phi) is the **angle of internal friction** or **shearing resistance**. The angle ϕ is not necessarily the slope angle but is the angle of internal friction within the slope mass and represents the angle of contact between the particles making up the soil or unconsolidated mass and the underlying surface. All unconsolidated materials tend to fail at angles less than the slope angle upon which they rest, loosely compacted materials failing at lower angles than compacted materials. The pressure of water in the soil voids, that is, the **pore water pressure**, ξ (xi), modifies the shear strength:

$$\tau_s = c + (\sigma - \xi) \tan \phi$$

This accounts for the common occurrence of slope failures after heavy rain, when pore water pressures are high and effective normal stresses $(\sigma - \xi)$ low. On 10 and 11 January 1999, a large portion of the upper part of Beachy Head, Sussex, England, collapsed (cf. p. 316). The rockfall appears to have resulted from increased pore pressures in the chalk following a wetter than normal year in 1998 and rain falling on most days in the fortnight before the fall.

The Mohr–Coulomb equation can be used to define the shear strength of a unit of rock resting on a failure plane and the susceptibility of that material to landsliding, providing the effects of fractures and joints are included. Whenever the stress applied to a soil or rock body is greater than the shear strength, the material will fail and move downslope. A scheme for defining the **intact rock strength** (the strength of rock excluding the effects of joints and fractures) has been devised. Intact strength is easily assessed using a Schmidt hammer, which measures the rebound of a known impact from a rock surface. Rock mass strength may be assessed using intact rock strength and other factors (weathering, joint spacing, joint orientations, joint width, joint continuity and infill, and groundwater outflow). Combining these factors gives a rock mass strength rating ranging from very strong, through strong, moderate, and weak, to very weak (see Selby 1980).

Soil behaviour

Materials are classed as rigid solids, elastic solids, plastics, or fluids. Each of these classes reacts differently to stress: they each have a characteristic relationship between the rate of deformation (strain rate) and the applied stress (shear stress) (Figure 3.4). Solids and liquids are easy to define. A perfect Newtonian **fluid** starts to deform immediately a stress is applied, the strain rate increasing linearly with the shear stress at a rate determined by the viscosity. **Solids** may have any amount of stress applied and remain rigid until the strength of the material is overstepped, at which point it will either deform or fracture depending on the rate at which the stress is applied. If a bar of hard toffee is suddenly struck, it behaves as a **rigid solid** and fractures. If gentle pressure is applied to it for some time, it behaves as an **elastic solid** and deforms reversibly before fracturing. Earth materials behave elastically when small stresses are applied to them. Perfect **plastic solids** resist deformation until the shear stress reaches a threshold value called the yield limit. Once beyond the yield stress, deformation of plastic bodies is unlimited and they do not revert to their original shape once the stress is withdrawn. **Liquids** include water and liquefied soils or sediments, that is, soil and sediments that behave as fluids.

An easy way of appreciating the rheology (response to stress) of different materials is to imagine a rubber ball, a clay ball, a glob of honey, and a cubic crystal of rock salt (cf. Selby 1982, 74). When dropped from the same height on to a hard floor, the elastic ball deforms on impact but quickly recovers its shape; the plastic clay sticks to the floor as a blob; the viscous honey spreads slowly over the floor; and the brittle rock salt crystal shatters and fragments are strewn over the floor.

Soil materials can behave as solids, elastic solids, plastics, or even fluids, in accordance with how much water they contain. In soils, clay content, along with the air and water content of voids, determines the mechanical behaviour. The **shrinkage limit** defines the point below which soils preserve a constant volume upon drying and behave as a solid. The **plastic limit** is minimum moisture content at which the soil can be moulded. The **liquid limit** is the point at which, owing to a high

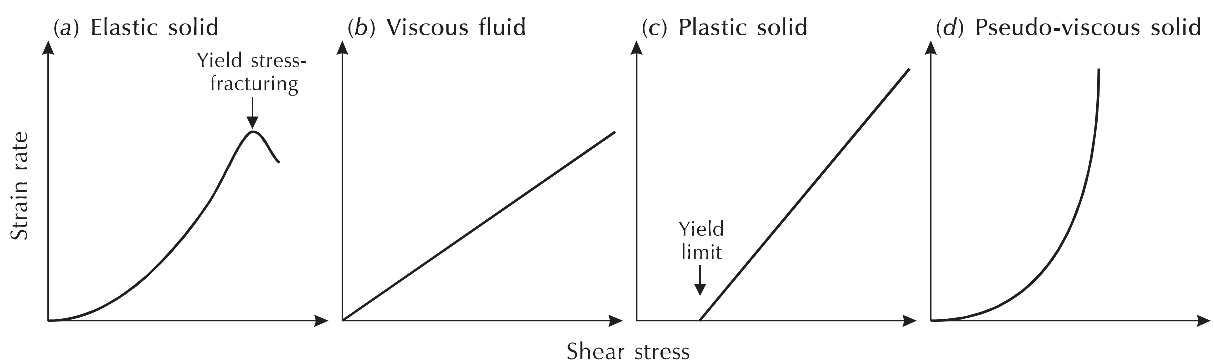


Figure 3.4 Stress–strain relationships in earth materials. (a) Elastic solids (rocks). (b) Viscous fluids (water and fluidized sediments). (c) Plastic solids (some soil materials). (d) Pseudo-viscous solids (ice).
Source: Adapted from Leopold *et al.* (1964, 31)

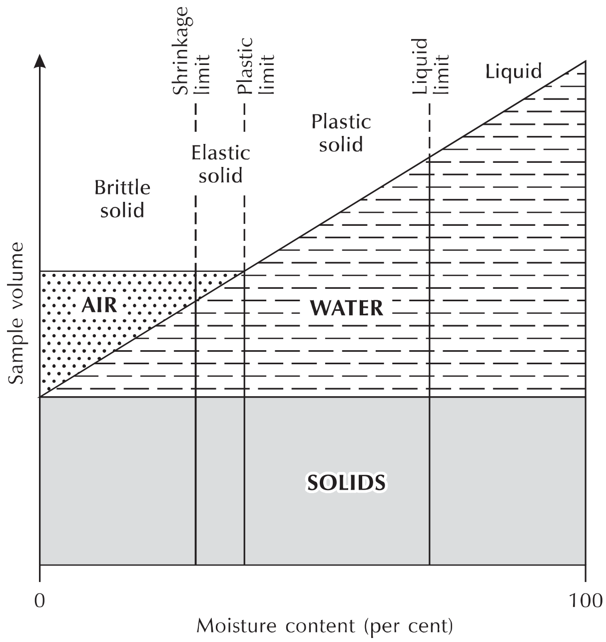


Figure 3.5 The composition of soil, ranging from air-filled pores, to water-filled pores, to a liquid. The Atterberg or soil limits are shown.
Source: Adapted from Selby (1982, 76)

moisture content, the soil becomes a suspension of particles in water and will flow under its own weight. The three limits separating different kinds of soil behaviour – shrinkage limit, plastic limit, and fluid limit – are known as **Atterberg limits**, after the Swedish soil scientist who first investigated them (Figure 3.5). The **plasticity index**, defined as the liquid limit minus the plastic limit, is an important indicator of potential slope instability. It shows the moisture range over which a soil will behave as a plastic. The higher the index, the less stable the slope.

Some soils, which are referred to as **quick clays** or **sensitive soils**, have a honeycomb structure that allows water content to go above the liquid limit. If such soils are subject to high shear stresses, perhaps owing to an earthquake or to burial, they may suddenly collapse, squeezing out water and turning the soil into a fluid. Quick clays are commonly associated with large and swift flows of slope materials. A violent shaking, as given by a seismic shock, may also liquefy a saturated mass of sand.

Mass movements

Mass movements may be classified in many ways. Table 3.2 summarizes a scheme recognizing six basic types and several subtypes, according to the chief mechanisms involved (creep, flow, slide, heave, fall, and subsidence) and the water content of the moving body (very low, low, moderate, high, very high, and extremely high):

1 **Rock creep** and **continuous creep** are the very slow plastic deformation of soil or rock. They result from stress applied by the weight of the soil or rock body and usually occur at depth, below the weathered mantle. They should not be confused with soil creep, which is a form of heave (see below).
2 **Flow** involves shear through the soil, rock, or snow and ice debris. The rate of flow is slow at the base of the flowing body and increases towards the surface. Most movement occurs as turbulent motion. Flows are classed as **avalanches** (the rapid downslope movement of earth, rock, ice, or snow), **debris flows**, **earthflows**, or **mudflows**, according to the predominant materials – snow and ice, rock debris, sandy material, or clay. Dry flows may also occur; water and ice flow. **Solifluction** and **gelifluction** – the downslope movement of saturated soil, the latter over permanently frozen subsoil – are the slowest flows. A **debris flow** is a fast-moving body of sediment particles with water or air or both that often has the consistency of wet cement. Debris flows occur as a series of surges lasting from a few seconds to several hours that move at 1 to 20 m/s. They may flow several kilometres beyond their source areas (Figure 3.6a). Some are powerful enough to destroy buildings and snap off trees that lie in their path. **Mudflows** triggered by water saturating the debris on the sides of volcanoes are called **lahars**. When Mount St Helens, USA, exploded on 18 May 1980 a huge debris avalanche mobilized a huge body of sediment into a remarkable lahar that ran 60 km from the volcano down the north and south forks of the Toutle River, damaging 300 km of road and 48 road bridges in the process.

Table 3.2 Mass movements and fluid movements

Main mechanism	*Water content*					
	Very low	*Low*	*Moderate*	*High*	*Very high*	*Extremely high*
Creep		Rock creep				
		Continuous creep				
Flow	Dry flow	Slow earthflow		Solifluction	Rapid earthflow	Mudflow
		Debris avalanche (struzstrom)		Gelifluction	Rainwash	Slush avalanche
		Snow avalanche (slab avalanche)		Debris flow	Sheet wash	Ice flow
		Sluff (small, loose snow avalanche)				Rill wash
						River flow
						Lake currents
Slide (translational)		Debris slide	Debris slide		Rapids (in part)	
		Earth slide	Earth slide		Ice sliding	
		Debris block slide	Debris block slide			
		Earth block slide	Earth block slide			
		Rockslide				
		Rock block slide				
Slide (rotational)		Rock slump	Debris slump			
			Earth slump			
Heave		Soil creep				
		Talus creep				
Fall		Rock fall				Waterfall
		Debris fall (topple)				Ice fall
		Earth fall (topple)				
Subsidence		Cavity collapse				
		Settlement				

Source: From Huggett (1997b, 196), partly adapted from Varnes (1978)

3 **Slides** are a widespread form of mass movement. They take place along clear-cut shear planes and are usually ten times longer than they are wide. Two subtypes are translational slides and rotational slides. **Translational slides** occur along planar shear planes and include debris slides, earth slides, earth block slides, rock slides, and rock block slides (Figure 3.6b). **Rotational slides**, also called **slumps**, occur along concave shear planes, normally under conditions of low to moderate water content, and are commonest on thick, uniform materials such as clays (Figure 3.6c; Plate 3.1). They include rock slumps, debris slumps, and earth slumps.

4 **Heave** is produced by alternating phases of expansion and contraction caused by heating and cooling, wetting and drying, and by the burrowing activities of animals. Material moves downslope during the cycles because expansion lifts material at right-angles to the slope but contraction drops it nearly vertically under the influence of gravity. Heave is classed as **soil**

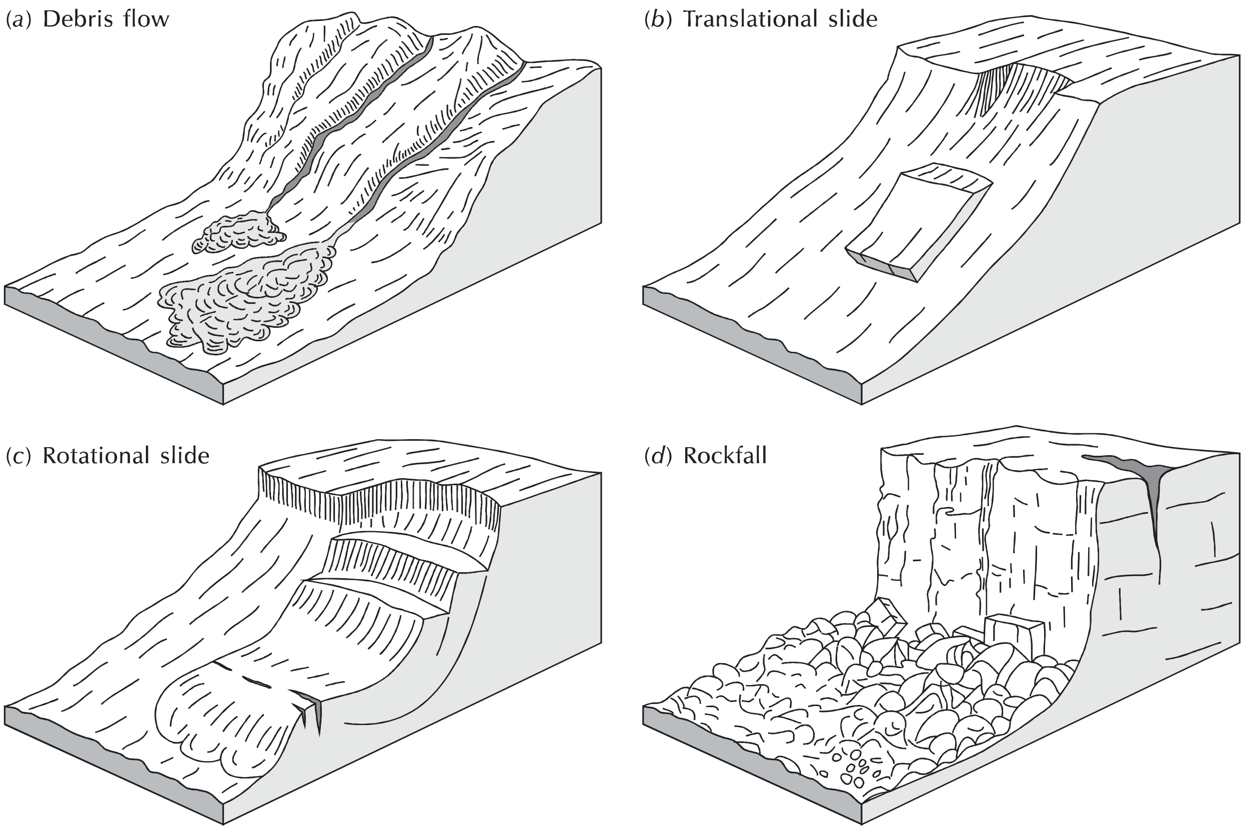


Figure 3.6 Some mass movements. (a) Flow. (b) Translational slide. (c) Rotational slide or slump. (d) Fall.

Plate 3.1 Shallow rotational landslide, Rockies foothills, Wyoming, USA.
(*Photograph by Tony Waltham Geophotos*)

creep (finer material) or **talus creep** (coarser material). **Frost creep** occurs when the expansion and contraction is brought about by freezing and thawing (pp. 77–8). **Terracettes** frequently occur on steep grassy slopes. Soil creep may produce them, although shallow landslides may be an important factor in their formation.

5 **Fall** is the downward movement of rock, or occasionally soil, through the air. Soil may topple from cohesive soil bodies, as in riverbanks. **Rock-falls** are more common, especially in landscapes with steep, towering rock slopes and cliffs (Figure 3.6d). Water and ice may also fall as waterfalls and icefalls. **Debris falls** and **earth falls**, also called debris and earth **topples**, occur, for example, along river banks.
6 **Subsidence** occurs in two ways: cavity collapse and settlement. First, in **cavity collapse**, rock or soil plummets into underground cavities, as in karst terrain (p. 198), in lava tubes, or in mining areas. In **settlement**, the ground surface is lowered progressively by compaction, often because of groundwater withdrawal or earthquake vibrations.

Gravity tectonics

Mass movements may occur on geological scales. Large rock bodies slide or spread under the influence of gravity to produce such large-scale features as **thrusts** and **nappes**. Most of the huge nappes in the European Alps and other intercontinental orogens are probably the product of massive **gravity slides**. Tectonic denudation is a term that describes the unloading of mountains by gravity sliding and spreading. The slides are slow, being only about 100 m/yr under optimal conditions (that is, over such layers as salt that offer little frictional resistance).

FLUVIAL PROCESSES

Flowing water

Figure 3.7 is a cartoon of the chief hydrological processes that influence the geomorphology of hillslopes and streams. Notice that water flows over and through landscapes in unconcentrated and concentrated forms.

Unconcentrated flow

Rainsplash results from raindrops striking rock and soil surfaces. An impacting raindrop compresses and spreads sideways. The spreading causes a shear on the rock or soil that may detach particles from the surface, usually particles less than 20 micrometres in diameter. If entrained by water from the original raindrop, the particles may rebound from the surface and travel in a parabolic curve, usually no more than a metre or so. Rainsplash releases particles for entrainment and subsequent transport by unconcentrated surface flow, which by itself may lack the power to dislodge and lift attached particles.

Unconcentrated surface flow (**overland flow**) occurs as inter-rill flow. **Inter-rill flow** is variously termed sheet flow, sheet wash, and slope wash. It involves a thin layer of moving water together with strands of deeper and faster-flowing water that diverge and converge around surface bulges causing erosion by soil detachment (largely the result of impacting raindrops) and sediment transfer. **Overland flow** is produced by two mechanisms:

1 **Hortonian overland flow** occurs when the rate at which rain is falling exceeds the rate at which it can percolate into the soil (the **infiltration rate**). Hortonian overland flow is more common on bare rock surfaces, and in deserts, where soils tend to be thin, bedrock outcrops common, vegetation scanty, and rainfall rates high. It can contribute large volumes of water to streamflow and cover large parts of an arid drainage basin, and is the basis of the 'partial area model' of streamflow generation.
2 **Saturation overland flow** or **seepage flow** occurs where the groundwater table sits at the ground surface. Some of the water feeding saturation overland flow is flow that has entered the hillside upslope and moved laterally through the soil as **throughflow**; this is called **return flow**. Rain falling directly on the hillslope may feed saturation overland flow.

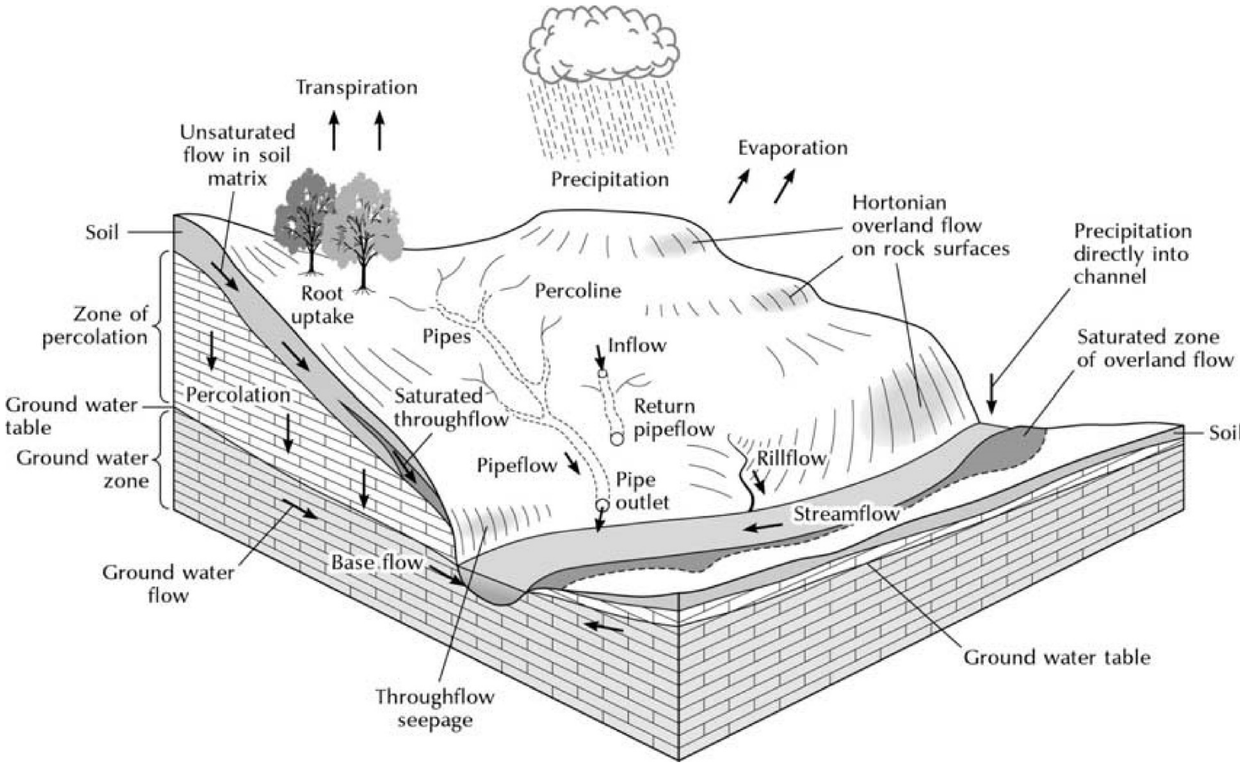


Figure 3.7 The chief hydrological processes that influence the geomorphology of hillslopes and streams. Water flows over and through landscapes in unconcentrated and concentrated forms.

Rill flow is deeper and speedier than inter-rill flow and is characteristically turbulent. It is a sporadic concentrated flow that grades into streamflow.

Subsurface flow

Flow within a rock or soil body may take place under unsaturated conditions, but faster **subsurface flow** is associated with localized soil saturation. Where the hydraulic conductivity of soil horizons decreases with depth, and especially when hardpans or clay-rich substrata are present in the soil, infiltrating water is deflected downslope as **throughflow**. Engineering hydrologists use the term **interflow** to refer to water arriving in the stream towards the end of a storm after having followed a deep subsurface route, typically through bedrock. **Baseflow** is water entering the stream from the water table or delayed interflow that keeps rivers in humid climates flowing during dry periods. Subsurface flow may take place as a slow movement through rock and soil pores, sometimes along distinct lines called **percolines**, or as a faster movement in cracks, soil pipes (**pipe flow**), and underground channels in caves.

Springs

Springs occur where the land surface and the water table cross. Whereas saturation overland flow is the seepage from a temporary saturation zone, springs arise where the water table is almost permanent. Once a spring starts to flow, it causes a dip in the water table that creates a pressure gradient in the aquifer. The pressure gradient then encourages water to move towards the spring. Several types of spring are recognized, including waste cover springs, contact springs, fault springs, artesian springs, karst springs, vauclusian springs, and geysers (Table 3.3).

Table 3.3 Springs

Type	*Occurrence*	*Example*
Waste cover	Dells and hollows where lower layers of soil or bedrock is impervious	Common on hillslopes in humid environments
Contact	Flat or gently dipping beds of differing perviousness or permeability at the contact of an aquifer and an aquiclude. Often occur as a spring line	Junction of Totternhoe Sands and underlying Chalk Marl, Cambridgeshire, England
Fault	Fault boundaries between pervious and impervious, or permeable and impermeable, rocks	Delphi, Greece
Artesian	Synclinal basin with an aquifer sandwiched between two aquicludes	Artois region of northern France
Karst	Karst landscapes	Orbe spring near Vallorbe, Switzerland
Vauclusian	U-shaped pipe in karst where water is under pressure and one end opens on to the land surface	Vaucluse, France; Blautopf near Blaubeuren, Germany
Thermal	Hot springs	Many in Yellowstone National Park, Wyoming, USA
Geyser	A thermal spring that spurts water into the air at regular intervals	Old Faithful, Yellowstone National Park

Streamflow

Rivers are natural streams of water that flow from higher to lower elevations across the land surface. Their continued existence relies upon a supply of water from overland flow, throughflow, interflow, baseflow, and precipitation falling directly into the river. **Channelized rivers** are streams structurally engineered to control floods, improve drainage, maintain navigation, and so on. In some lowland catchments of Europe, more than 95 per cent of river channels have been altered by channelization.

Water flowing in an open channel (**open channel flow**) is subject to gravitational and frictional forces. Gravity impels the water downslope, while friction from within the water body (viscosity) and between the flowing water and the channel surface resists movement. **Viscosity** arises through cohesion and collisions between molecules (**molecular** or **dynamic viscosity**) and the interchange of water adjacent to zones of flow within eddies (**eddy viscosity**).

Water flow may be turbulent or laminar. In **laminar flow**, thin layers of water 'slide' over each other, with resistance to flow arising from molecular viscosity (Figure 3.8a). In **turbulent flow**, which is the predominant type of flow in stream channels, the chaotic flow-velocity fluctuations are superimposed on the main forward flow, and resistance is contributed by molecular viscosity and eddy viscosity. In most channels, a thin layer or laminar flow near the stream bed is surmounted by a much thicker zone of turbulent flow (Figure 3.8b). Mean flow velocity, molecular viscosity, fluid density, and the size of the flow section determine the type of flow. The size of the flow section may be measured as either the depth of flow or as the hydraulic radius. The **hydraulic radius**, R, is the cross-sectional area of flow divided by the wetted perimeter, P, which is the length of the boundary along which water is in contact with the channel (Figure 3.9). In broad, shallow channels, the flow depth can approximate the hydraulic radius. The **Reynolds number**, R_e,

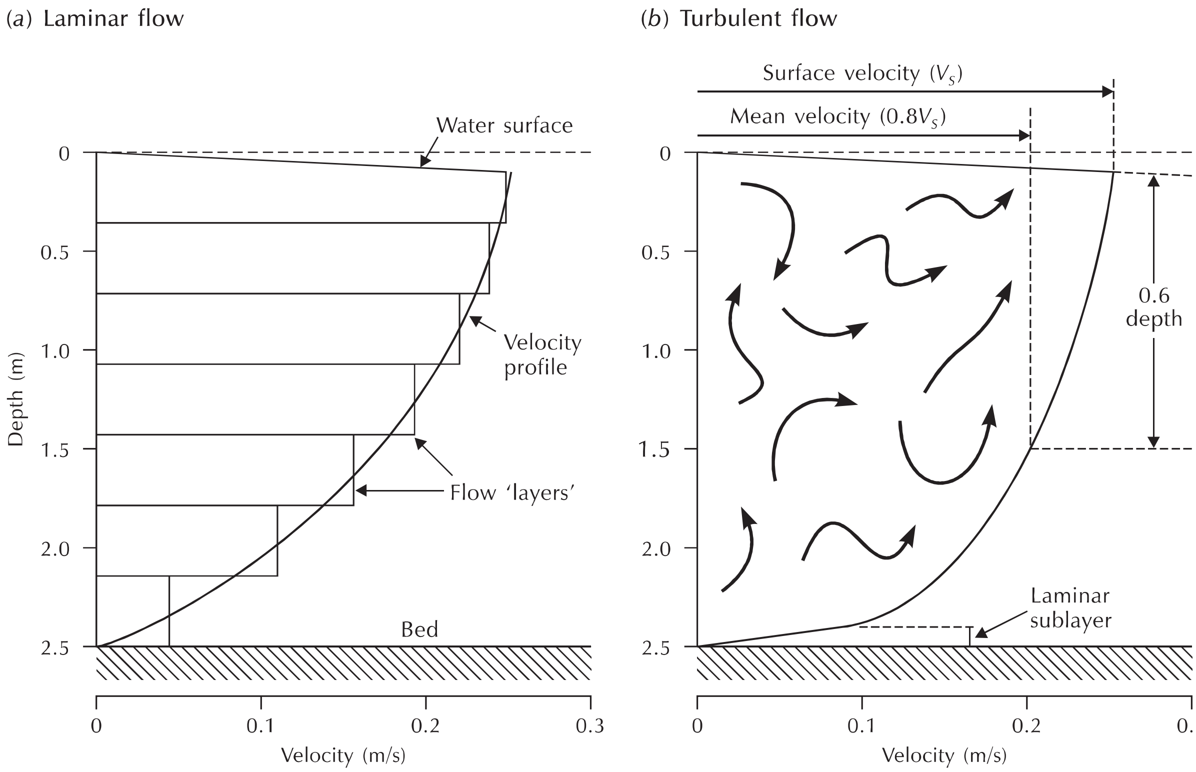


Figure 3.8 Velocity profiles of (a) laminar and (b) turbulent flow in a river.

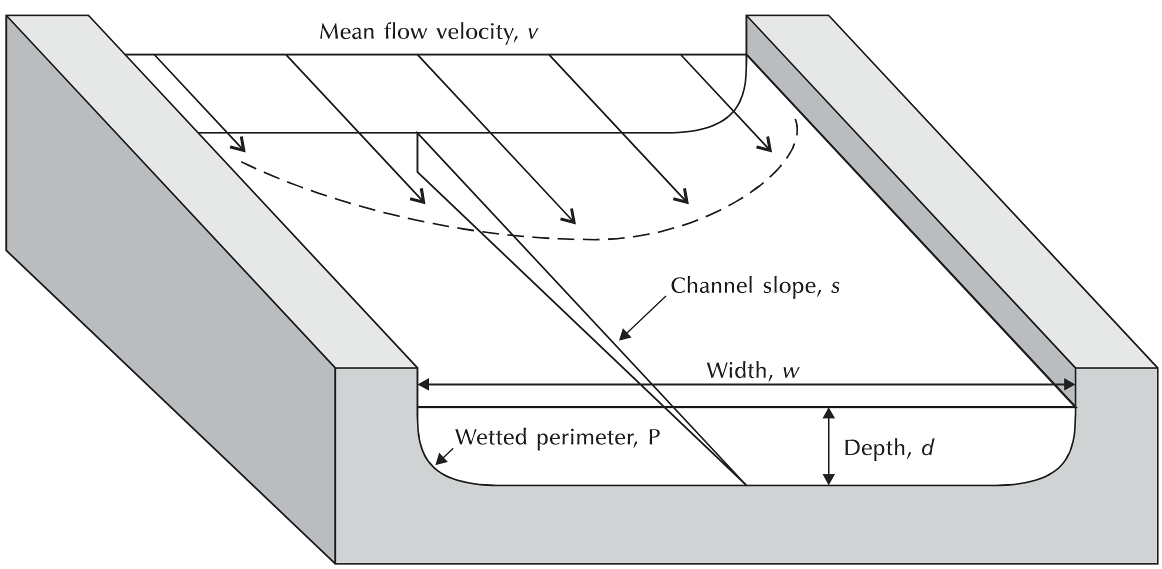


Figure 3.9 Variables used in describing streamflow.

Box 3.4

REYNOLDS AND FROUDE NUMBERS

Reynolds number is a dimensionless number that includes the effects of the flow characteristics, velocity, and depth, and the fluid density and viscosity. It may be calculated by multiplying the mean flow velocity, v, and hydraulic radius, R, and dividing by the kinematic viscosity, ν (nu), which represents the ratio between molecular viscosity, μ (mu), and the fluid density, σ (rho):

$$R_e = \frac{\rho \nu R}{\mu}$$

For stream channels at moderate temperatures, the maximum Reynolds number at which laminar flow is sustained is about 500. Above values of about 2,000, flow is turbulent, and between 500 and 2,000 laminar and turbulent flow are both present.

The **Froude number** is defined by the square root of the ratio of the inertia force to the gravity force, or the ratio of the flow velocity to the velocity of a small gravity wave (a wave propagated by, say, a tossed pebble) in still water. The Froude number is usually computed as:

$$F = \frac{v}{\sqrt{gd}}$$

where v is the flow velocity, g is the acceleration of gravity, d is the depth of flow, and $\sqrt{gd}$ is the velocity of the gravity waves. When $F < 1$ (but more than zero) the wave velocity is greater than the mean flow velocity and the flow is known as **subcritical** or **tranquil** or **streaming**. Under these conditions, ripples propagated by a pebble dropped into a stream create an egg-shaped wave that moves out in all directions from the point of impact. When $F = 1$ flow is critical, and when $F > 1$ it is **supercritical** or **rapid** or **shooting**. These different types of flow occur because changes in discharge can be accompanied by changes in depth and velocity of flow. In other words, a given discharge can be transmitted along a stream channel either as a deep, slow-moving, subcritical flow or else as a shallow, rapid, supercritical flow. In natural channels, mean Froude numbers are not usually higher than 0.5 and supercritical flows are only temporary, since the large energy losses that occur with this type of flow promote bulk erosion and channel enlargement. This erosion results in a lowering of flow velocity and a consequential reduction in the Froude number of the flow through negative feedback. For a fixed velocity, streaming flow may occur in deeper sections of the channel and shooting flow in shallower sections.

named after English scientist and engineer Osborne Reynolds, may be used to predict the type of flow (laminar or turbulent) in a stream (Box 3.4).

In natural channels, irregularities on the channel bed induce variations in the depth of flow, so propagating ripples or waves that exert a weight or gravity force. The **Froude number**, F, of the flow, named after the English engineer and naval architect William Froude, can be used to distinguish different states of flow – **subcritical flow** and **critical flow** (Box 3.4). **Plunging flow** is a third kind of turbulent flow. It occurs at a waterfall, when water plunges in free fall over very steep, often vertical or overhanging rocks. The water falls as a coherent mass or as individual water strands or, if the falls are very high and the discharge low, as a mist resulting from the water dissolving into droplets.

Flow velocity controls the switch between subcritical and supercritical flow. A **hydraulic jump** is a sudden change from supercritical to subcritical flow. It produces a stationary wave and an increase in water

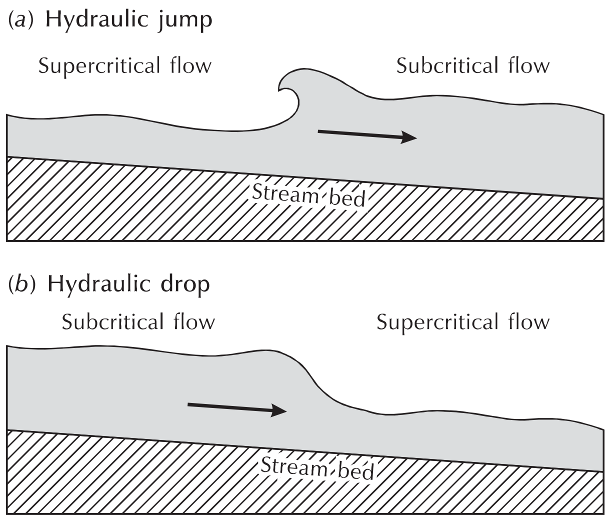


Figure 3.10 (a) Hydraulic jump. (b) Hydraulic drop.

depth (Figure 3.10a). A **hydraulic drop** marks a change from subcritical to supercritical flow and is accompanied by a reduction in water depth (Figure 3.10b). These abrupt changes in flow regimes may happen where there is a sudden change in channel bed form, a situation rife in mountain streams where there are usually large obstructions such as boulders.

Flow velocity in streams is affected by the slope gradient, bed roughness, and cross-sectional form of the channel. It is very time-consuming to measure streamflow velocity directly, and empirical equations have been devised to estimate mean flow velocities from readily measured channel properties. The **Chézy equation**, named after the eighteenth-century French hydraulic engineer Antoine de Chézy, estimates velocity in terms of the hydraulic radius and channel gradient, and a coefficient expressing the gravitational and frictional forces acting upon the water. It defines mean flow velocity, $\overline{v}$, as:

$$\overline{v} = C\sqrt{Rs}$$

where R is the hydraulic radius, s is the channel gradient, and C is the Chézy coefficient representing gravitational and frictional forces. The **Manning equation**, which was devised by the American hydraulic engineer Robert Manning at the end of the nineteenth century, is a more commonly used formula for estimating flow velocity:

$$\overline{v} = \frac{R^{2/3}s^{1/2}}{n}$$

where R is the hydraulic radius, s the channel gradient, and n the **Manning roughness coefficient**, which is an index of bed roughness and is usually estimated from standard tables or by comparison with photographs of channels of known roughness. Manning's formula can be useful in estimating the discharge in flood conditions. The height of the water can be determined from debris stranded in trees and high on the bank. Only the channel cross-section and the slope need measuring.

Fluvial erosion and transport

Streams are powerful geomorphic agents capable of eroding, carrying, and depositing sediment. **Stream power** is the capacity of a stream to do work. It may be expressed as:

$$\Omega = \rho g Q s$$

where Ω (omega) is stream power per unit length of stream channel, ρ (rho) is water density, Q is stream discharge, and s is the channel slope. It defines the rate at which potential energy, which is the product of the weight of water, mg (mass, m, times gravitational acceleration, g), and its height above a given datum, h, is expended per unit length of channel. In other words, stream power is the rate at which a stream works to transport sediment, overcome frictional resistance, and generate heat. It increases with increasing discharge and increasing channel slope.

Stream load

All the material carried by a stream is its **load**. The **total load** consists of the dissolved load (solutes), the suspended load (grains small enough to be suspended in the water), and the bed load (grains too large to be suspended for very long under normal flow conditions). In detail, the three components of stream load are as follows:

1 The **dissolved load** or **solute load** comprises ions and molecules derived from chemical weathering plus some dissolved organic substances. Its composition depends upon several environmental factors, including climate, geology, topography, and vegetation. Rivers fed by water that has passed though swamps, bogs, and marshes are especially rich in dissolved organic substances. River waters draining large basins tend to have a similar chemical composition, with bicarbonate, sulphate, chloride, calcium, and sodium being the dominant ions (but see p. 43 for continental differences). Water in smaller streams is more likely to mirror the composition of the underlying rocks.
2 The **suspended load** consists of solid particles, mostly silts and clays, that are small enough and light enough to be supported by turbulence and vortices in the water. Sand is lifted by strong currents, and small gravel can be suspended for a short while during floods. The suspended load reduces the inner turbulence of the stream water, so diminishing frictional losses and making the stream more efficient. Most of the suspended load is carried near the stream bed, and the concentrations become lower in moving towards the water surface.
3 The **bed load** or **traction load** consists of gravel, cobbles, and boulders, which are rolled or dragged along the channel bed by traction. If the current is very strong, they may be bounced along in short jumps by saltation. Sand may be part of the bed load or part of the suspended load, depending on the flow conditions. The bed load moves more slowly than the water flows as the grains are moved fitfully. The particles may move singly or in groups by rolling and sliding. Once in motion, large grains move more easily and faster than small ones, and rounder particles move more readily than flat or angular ones. A stream's **competence** is defined as the biggest size of grain that a stream can move in traction as bed load. Its **capacity** is defined as the maximum amount of debris that it can carry in traction as bed load.

In addition to these three loads, the suspended load and the bed load are sometimes collectively called the **solid-debris load** or the **particulate load**. And the **wash load**, a term used by some hydrologists, refers to that part of the sediment load comprising grains finer than those on the channel bed. It consists of very small clay-sized particles that stay in more or less permanent suspension.

Stream erosion and transport

Streams may attack their channels and beds by corrosion, corrasion, and cavitation. **Corrosion** is the chemical weathering of bed and bank materials in contact with the stream water. **Corrasion** or **abrasion** is the wearing away of surfaces over which the water flows by the impact or grinding action of particles moving with the water body. **Evorsion** is a form of corrasion in which the sheer force of water smashes bedrock without the aid of particles. In alluvial channels, **hydraulicking** is the removal of loose material by the impact of water alone. **Cavitation** occurs only when flow velocities are high, as at the bottom of waterfalls, in rapids, and in some artificial conduits. It involves shockwaves released by imploding bubbles, which are produced by pressure changes in fast-flowing streams, smashing into the channel walls, hammer-like, and causing rapid erosion. The three main erosive processes are abetted by vortices that may develop in the stream and that may suck material from the streambed.

Streams may erode their channels downwards or sideways. **Vertical erosion** in an alluvial channel bed (a bed formed in fluvial sediments) takes place when there is a net removal of sands and gravels. In bedrock channels (channels cut into bedrock), vertical erosion is caused by the channel's bed load abrading the bed. **Lateral erosion** occurs when the channel banks are worn away, usually by being undercut, which leads to slumping and bank collapse.

The ability of flowing water to erode and transport rocks and sediment is a function of a stream's kinetic energy (the energy of motion). Kinetic energy, E_k, is half the product of mass and velocity, so for a stream it may be defined as

$$E_k = mv^2/2$$

where m is the mass of water and v is the flow velocity. If Chézy's equation (p. 71) is substituted for velocity, the equation reads

$$E_k = (mCRs)/2$$

This equation shows that kinetic energy in a stream is directly proportional to the product of the hydraulic radius, R (which is virtually the same as depth in large rivers), and the stream gradient, s. In short, the deeper and faster a stream, the greater its kinetic energy and the larger its potential to erode. The equation also conforms to the **DuBoys equation** defining the shear stress or tractive force, τ (tau), on a channel bed:

$$\tau = \gamma ds$$

where γ (gamma) is the specific weight of the water (g/cm^3), d is water depth (cm), and s is the stream gradient expressed as a tangent of the slope angle. A stream's ability to set a pebble in motion – its **competence** – is largely determined by the product of depth and slope (or the square of its velocity). It can move a pebble of mass m when the shear force it creates is equal to or exceeds the critical shear force necessary for the movement of the pebble, which is determined by the mass, shape, and position of the pebble in relation to the current. The pebbles in gravel bars often develop an imbricated structure (overlapping like tiles on a roof), which is particularly resistant to erosion. In an imbricated structure, the pebbles have their long axes lying across the flow direction and their second-longest axes aligned parallel to the flow direction and angled down upstream. Consequently, each pebble is protected by its neighbouring upstream pebble. Only if a high discharge occurs are the pebbles set in motion again.

A series of experiments enabled Filip Hjulstrøm (1935) to establish relationships between a stream's flow velocity and its ability to erode and transport grains of a particular size. The relationships, which are conveniently expressed in the oft-reproduced **Hjulstrøm diagram** (Figure 3.11), cover a wide range of grain sizes and flow velocities. The upper curve is a band showing the critical velocities at which grains of a given size start to erode. The curve is a band rather than a single line because the critical velocity depends partly on the position of the grains and the way that they lie on the bed. Notice that medium sand (0.25–0.5 mm) is eroded at the lowest velocities. Clay and silt particles, even though they are smaller than sand particles, require a higher velocity for erosion to occur because they lie within the bottom zone of laminar flow and, in the case of clay particles, because of the cohesive forces holding them together. The lower curve in the Hjulstrøm diagram shows the velocity at which particles already in motion cannot be transported further and fall to the channel bed. This is called the fall velocity. It depends not just on grain size but on density and shape, too, as well as on the viscosity and density of the water. Interestingly, because the viscosity and density of the water change with the amount of sediment the stream carries, the relationship between flow velocity and deposition is complicated. As the flow velocity reduces, so the coarser grains start to fall out, while the finer grains remain in motion. The result is differential settling and **sediment sorting**. Clay and silt particles stay in suspension at velocities of 1–2 cm/s, which explains why suspended load deposits are not dumped on streambeds. The region between the lower curve and the upper band defines the velocities at which particles of different sizes are transported. The wider the gap between the upper and lower lines, the more continuous the transport. Notice that the gap for particles larger than 2 mm is small. In consequence, a piece of gravel eroded at just above the critical velocity will be deposited as soon as it arrives in a region of slightly lower velocity, which is likely to lie near the point of erosion. As a rule of thumb, the flow velocity at which erosion starts for grains larger than 0.5 mm is roughly proportional to the square root of the grain size. Or, to put it another way, the maximum grain size eroded is proportional to the square of the flow velocity.

The Hjulstrøm diagram applies only to erosion, transport, and deposition in alluvial channels. In bedrock channels, the bed load abrades the rock floor and causes vertical erosion. Where a stationary eddy forms, a small

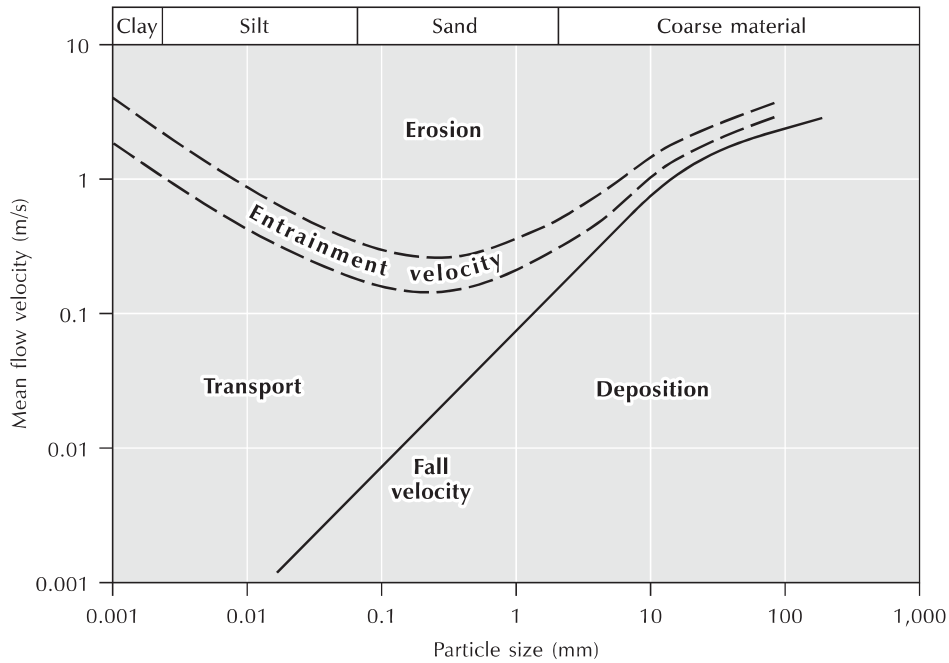


Figure 3.11 The Hjulstrøm diagram showing the water velocity at which entrainment and deposition occur for particles of a given size in well-sorted sediments.
Source: Adapted from Hjulstrøm (1935)

hollow is ground out that may eventually deepen to produce a **pothole**.

Channel initiation

Stream channels can be created on a newly exposed surface or develop by the expansion of an existing channel network. Their formation depends upon water flowing over a slope becoming sufficiently concentrated for channel incision to occur. Once formed, a channel may grow to form a permanent feature.

Robert E. Horton (1945) was the first to formalize the importance of topography to hillslope hydrology by proposing that a **critical hillslope length** was required to generate a channel (cf. p. 66). The critical length was identified as that required to generate a boundary shear stress of Hortonian overland flow sufficient to overcome the surface resistance and result in scour. In Horton's model, before overland flow is able to erode the soil, it has to reach a critical depth at which the eroding stress of the flow exceeds the shear resistance of the soil surface (Figure 3.12). Horton proposed that a '**belt of no erosion**' is present on the upper part of slopes because here the flow depth is not sufficient to cause erosion. However, subsequent work has demonstrated that some surface wash is possible even on slope crests, although here it does not lead to rill development because the rate of incision is slow and incipient rills are filled by rainsplash.

Further studies have demonstrated that a range of relationships between channel network properties and topography exist, although the physical processes driving these are not as well understood. In semi-arid and arid environments, the Hortonian overland-flow model provides a reasonable framework for explaining channel initiation, but it does not for humid regions. In humid regions, channel initiation is more related to the location of surface and subsurface flow convergence, usually in slope concavities and adjacent to existing drainage lines, than to a critical distance of overland flow. Rills can develop as a result of a sudden outburst of subsurface flow at the surface close to the base of a slope. So channel

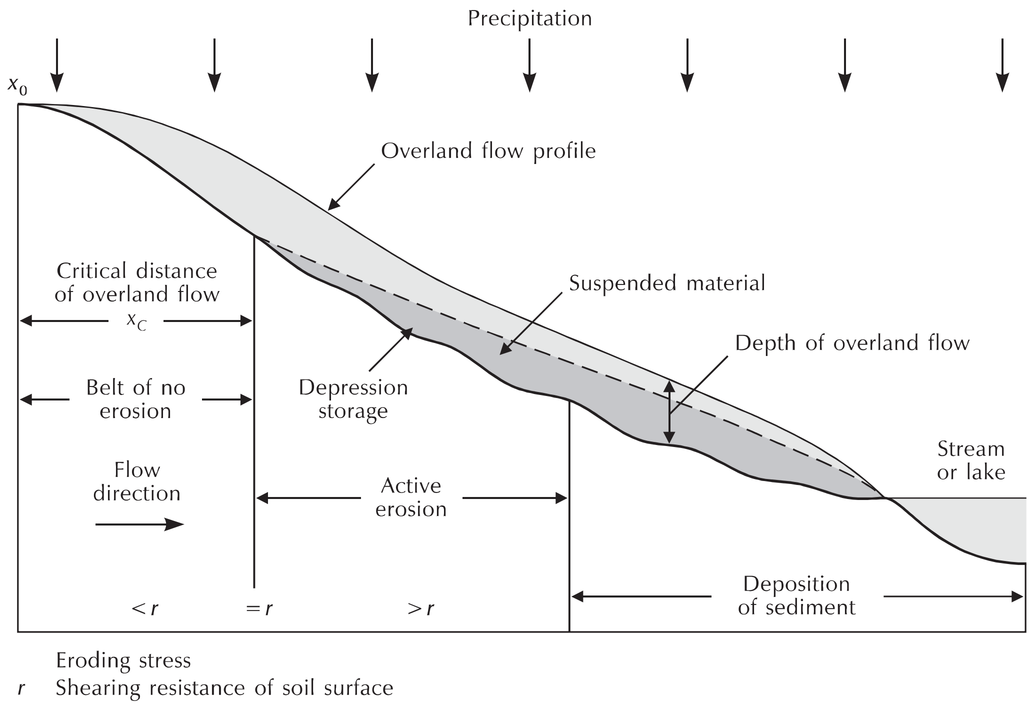


Figure 3.12 Horton's model of overland flow production.
Source: Adapted from Horton (1945)

development in humid regions is very likely to occur where subsurface **pipes** are present. Pipe networks can help initiate channel development, either through roof collapse or by the concentration of runoff and erosion downslope of pipe outlets. Piping can also be important in semi-arid regions. Channel initiation may also take place where slope wash and similar mass movements dominate soil creep and creep-like processes (e.g. Smith and Bretherton 1972; Tarboton *et al.* 1992).

Fluvial deposition

Rivers may deposit material anywhere along their course, but they mainly deposit material in valley bottoms where gradients are low, at places where gradients change suddenly, or where channelled flow diverges, with a reduction in depth and velocity. The Hjulstrøm diagram (p. 74) defines the approximate conditions under which solid-load particles are deposited upon the stream bed. Four types of fluvial deposit are recognized: **channel deposits**, **channel margin deposits**, **overbank flood plain deposits**, and **valley margin deposits** (Table 3.4).

When studying stream deposition, it is useful to take the broad perspective of erosion and deposition within drainage basins. Stream erosion and deposition take place during flood events. As discharge increases during a flood, so erosion rates rise and the stream bed is scoured. As the flood abates, sediment is redeposited over days or weeks. Nothing much then happens until the next flood. Such **scour-and-fill cycles** shift sediment along the streambed. Scour-and-fill and channel deposits are found in most streams. Some streams actively accumulate sediment along much of their courses, and many streams deposit material in broad expanses in the lower reaches but not in their upper reaches. **Alluviation** is large-scale deposition affecting much of a stream system. It results from fill preponderating scour for long periods of time. As a general rule, scour and erosion dominate upstream channels, and fill and deposition dominate downstream channels. This pattern arises from steeper stream gradients, smaller hydraulic radii,

Table 3.4 Classification of valley sediments

Type of deposit	*Description*
Channel deposits	
Transitory channel deposits	Resting bed-load. Part may be preserved in more durable channel fills or lateral accretions
Lag deposits	Sequestrations of larger or heavier particles. Persist longer than transitory channel deposits
Channel fills	Sediment accumulated in abandoned or aggrading channel segments. Range from coarse bed-load to fine-grained oxbow lake deposits
Channel margin deposits	
Lateral accretion deposits	Point bars and marginal bars preserved by channel shifting and added to the overbank floodplain
Overbank floodplain deposits	
Vertical accretion deposits	Fine-grained sediment deposited from the load suspended in overbank flood-water. Includes natural levees and backswamp deposits
Splays	Local accumulations of bed-load materials spread from channel on to bordering floodplains
Valley margin deposits	
Colluvium	Deposits derived mainly from unconcentrated slope wash and soil creep on valley sides bordering floodplains
Mass movement deposits	Debris from earthflow, debris avalanches, and landslides, commonly intermixed with marginal colluvium. Mudflows normally follow channels but may spill over the channel bank

Source: Adapted from Benedict *et al.* (1971)

and rougher channels upstream promoting erosion; and shallower gradients, larger hydraulic radii, and smoother channels downstream promoting deposition. In addition, flat, low-lying land bordering a stream that forms a suitable platform for deposition is more common at downstream sites.

Alluviation may be studied by calculating **sediment budgets** for alluvial or valley storage in a drainage basin. The change in storage during a time interval is the difference between the sediment gains and the sediment losses. Where gains exceed losses, storage increases with a resulting **aggradation** of channels or floodplains or both. Where losses exceed gains, channels and floodplains are eroded (**degraded**). It is feasible that gains counterbalance losses to produce a steady state. This condition is surprisingly rare, however. Usually, valley storage and fluxes conform to one of four common patterns under natural conditions (Trimble 1995): a quasi-steady-state typical of humid regions, vertical accretion of channels and aggradation of floodplains, valley trenching (arroyo cutting), episodic gains and losses in mountain and arid streams (Figure 3.13).

GLACIAL PROCESSES

Ice, snow, and frost are solid forms of water. Each is a powerful geomorphic agent. It is convenient to discuss

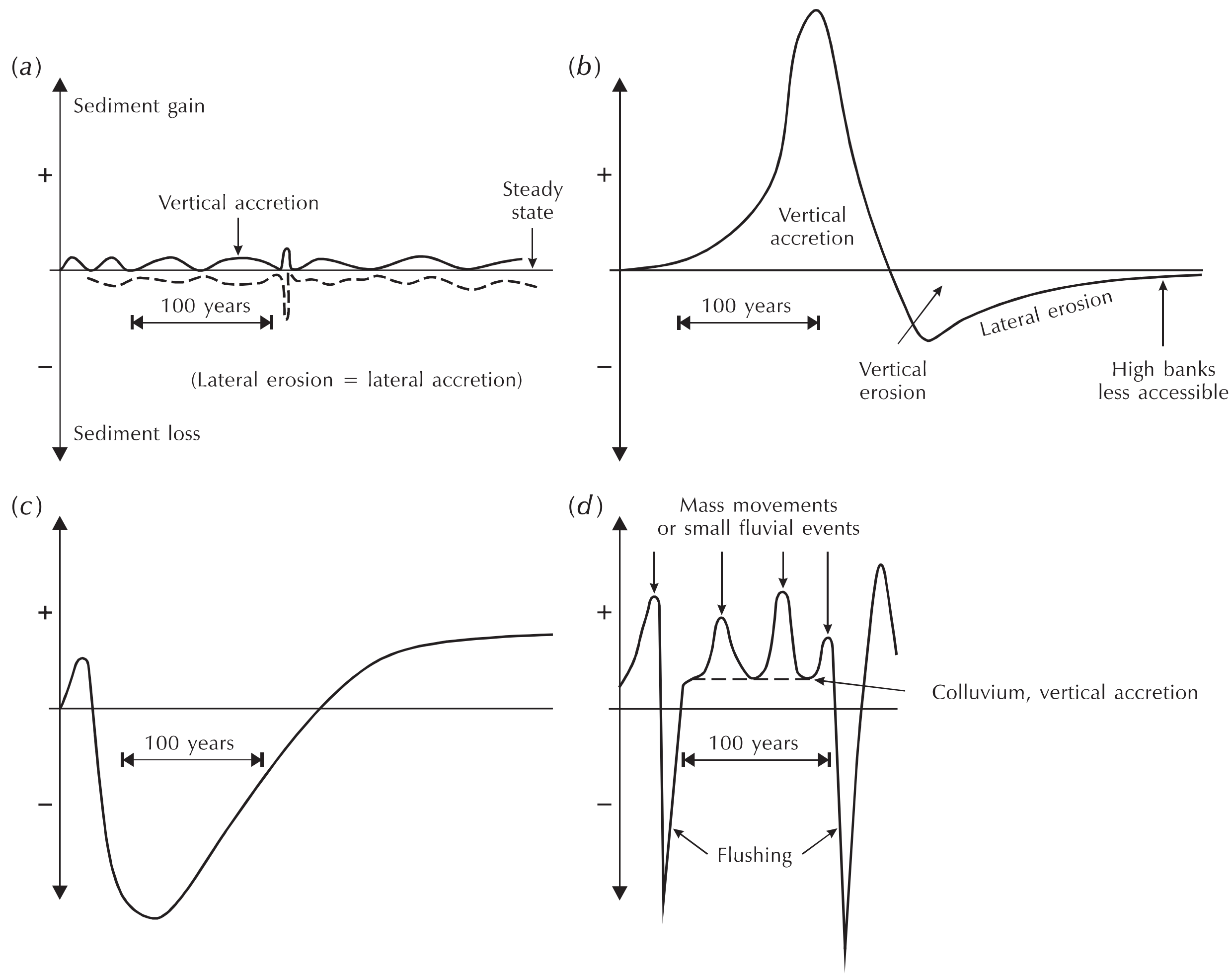


Figure 3.13 Four common patterns of valley sediment storage and flux under natural conditions. (a) Quasi-steady-state typical of humid regions. (b) Great sediment influx with later amelioration producing vertical accretion of channels and aggradation of floodplains. (c) Valley trenching (arroyo cutting). (d) High-energy instability seen as episodic gains and losses in mountain and arid streams.
Source: Adapted from Trimble (1995)

frost and snow processes separately from processes associated with flowing ice in glaciers.

Frost and snow processes

The freezing of water in rock, soil, and sediment gives rise to several processes – frost shattering, heaving and thrusting, and cracking – that are intense in the periglacial zone. Water in the ground may freeze *in situ* within voids, or it may migrate through the voids (towards areas where temperatures are sub-zero) to form discrete masses of segregated ice. **Segregated ice** is common in sediments dominated by intermediate grain sizes, such as silt. Coarse sediments, such as gravel, are too permeable and very fine-grained sediments, such as clay, too impermeable and have too high a suction potential (the force with which water is held in the soil body) for segregation to occur. Frost action is crucially determined by the occurrence of freeze–thaw cycles at the ground surface. Freeze–thaw cycles are mainly determined by air temperature fluctuations, but they are modulated by the thermal properties of the ground-surface materials, vegetation cover, and snow cover.

Frost weathering and shattering

Frost weathering was covered in an earlier section (p. 51). Many periglacial landscapes are carpeted by angular rock debris, the origin of which is traditionally attributed to **frost shattering**. However, frost shattering requires freeze–thaw cycles and a supply of water. Field investigations, which admittedly are not yet large in number, indicate that such conditions may not be as common as one might imagine. Other processes, such as hydration shattering and salt weathering (in arid and coastal sites), may play a role in rock disintegration. It is also possible that, especially in lower-latitude glacial environments, the pervasive angular rock debris is a relict of Pleistocene climates, which were more favourable to frost shattering.

Frost heaving and thrusting

Ice formation causes **frost heaving**, which is a vertical movement of material, and **frost thrusting**, which is a horizontal movement of material. Heaving and thrusting normally occur together, though heaving is probably predominant because the pressure created by volume expansion of ice acts parallel to the direction of the maximum temperature gradient, which normally lies at right-angles to the ground surface. Surface stones may be lifted when needle ice forms. Needle ice or pipkrake forms from ice crystals that extend upwards to a maximum of about 30 mm (cf. Table 11.1). Frost heaving in the active layer seems to result from three processes: ice-lens growth as downward freezing progresses; ice-lens growth near the bottom of the active layer caused by upward freezing from the permafrost layer; and the progressive freezing of pore water as the active layer cools below freezing point. Frost heaving displaces sediments and appears to occasion the differential vertical movement of sedimentary particles of different sizes. In particular, the upward passage of stones in periglacial environments is a widely observed phenomenon. The mechanisms by which this process arises are debatable. Two groups of hypotheses have emerged: the frost-pull hypotheses and the frost-push hypotheses. In essence, **frost-pull** involves all soil materials rising with ground expansion on freezing, followed by the collapse of fine material on thawing while larger stones are still supported on ice. When the ice eventually melts, the fine materials support the stones. **Frost-push** consists of flowing water tending to collect beneath a stone and on freezing lifting it. On melting, finer particles fall into the void and the stone falls back on top of them. The frost-push mechanism is known to work under laboratory conditions but applies to stones near the surface. The frost-pull mechanism is in all likelihood the more important under natural circumstances.

Mass displacement

Frost action may cause local vertical and horizontal movements of material within soils. Such **mass displacement** may arise from cryostatic pressures within pockets of unfrozen soil caught between the permafrost table and the freezing front. However, differential heating resulting from annual freezing and thawing would lead to a similar effect. It is possible that, towards the feet of slopes, positive pore-water pressures would bring about mass displacement to form periglacial **involutions** in the active layer. Periglacial involutions consist of interpenetrating layers of sediment that originally lay flat.

Frost cracking

At sub-zero temperatures, the ground may crack by thermal contraction, a process called **frost cracking**. The polygonal fracture patterns so prevalent in periglacial environments largely result from this mechanism, though similar systems of cracks are made by drying out (**desiccation cracking**) and by differential heaving (**dilation cracking**).

Frost creep and gelifluction

Most kinds of mass movement occur in periglacial environments, but **frost creep** and **solifluction** are of paramount significance (p. 66). **Solifluction** commonly