Strahler, A.N. (1952) Dynamic basis of geomorphology, *Geological Society of America Bulletin* 63, 923–938.
——(1980) Systems theory in physical geography, *Physical Geography* 1, 1–27.
——(1992) Quantitative/dynamic geomorphology at Columbia (1945–1960): a retrospective, *Progress in Physical Geography* 16, 65–84.
Von Bertalanffy, L. (1950) The theory of open systems in physics and biology, *Science* 111, 23–28.

SEE ALSO: cyclic time; dynamic equilibrium; force and resistance concept; graded time

OLAV SLAYMAKER

E

EFFECTIVE STRESS

The difference between total stress (σ) and pore pressure in a material (u), responsible for mobilizing internal friction. Effective stress (σ') is one of the two components of internal stress within a material, alongside pore pressure, and measures the distribution of load carried by the soil over a specific area. The principle of effective stress was developed by Karl Terzaghi between 1923–1936, and is a fundamental theory in soil mechanics. Changes in stress, such as distortion, compression and shearing resistance changes, are due to variations in effective stress. As effective stress values increase, the soil or rock becomes more consolidated, exhibiting a maximum value at complete consolidation and before shear failure. Thus, it is the effective stress that causes important changes in material strength, volume and shape. Long-term SLOPE STABILITY analysis often incorporates effective stress analysis (inclusive of internal stresses), rather than total stress analysis (short-term slope instability).

Further reading

Lade, P.V. and De Boer, R. (1997) The concept of effective stress for soil, concrete, and rock, *Géotechnique* 47(1), 61–78.
Clayton, C.R.I., Muller-Steinhagen, H. and Powrie, W. (1995) Terzaghi's theory of consolidation, and the discovery of effective stress, *Proceedings International Conference on Engineering (Geotechnical Engineering)* 113(4), 191–205.

STEVE WARD

EL NIÑO EFFECTS

Climate oscillations occur at many timescales. For example, in the tropics a sub-annual, intra-seasonal 40–60-day period Madden–Julian oscillation has been identified. At a slightly longer timescale there is a 2 to 2.5 year oscillation in the equatorial jet in the lower stratosphere, and this is called the Quasi-Biennial Oscillation (QBO). Every three to seven years a two or so year-long event occurs which is called the El Niño Southern Oscillation (ENSO). Decadal and interdecadal variability is evident in the North Atlantic Oscillation (NAO). At even longer timescales there are such important phenomena as the Dansgaard–Oeschger Cycles, Bond Cycles and Heinrich events, which occur at century to millennial scales. El Niño is the term used to describe an extensive warming of the upper ocean in the tropical eastern Pacific lasting up to a year or even more. The negative or cooling phase of El Niño is called *La Niña*. El Niño events are linked with a change in atmospheric pressure known as the Southern Oscillation (SO). Because the SO and El Niño are so closely linked, they are often known collectively as the El Niño/Southern Oscillation or ENSO. The system oscillates between warm to neutral (or cold) conditions every three to four years.

Precipitation and temperature anomalies appear to characterize all El Niño warm episodes. These include:

- The eastward shift of thunderstorm activity from Indonesia to the central Pacific usually results in abnormally dry conditions over north Australia, Indonesia and the Philippines.
- Drier-than-normal conditions are also usually observed over south-eastern Africa and north Brazil.
- During the northern summer season, the Indian monsoon rainfall tends to be less than normal, especially in the north-west of the subcontinent.

- Wetter-than-normal conditions are usual along the west coast of tropical south America, and at subtropical latitudes of North America (the Gulf Coast) and South America (south Brazil to central Argentina).
- El Niño conditions are thought to suppress the development of tropical storms and hurricanes in the Atlantic but to increase the numbers of tropical storms over the eastern and central Pacific Ocean.

In the twentieth century there were around twenty-five warm events of differing strengths, with that of 1997/8 being seen as especially strong. ENSO was relatively quiescent from the 1920s to 1940s.

Severe El Niños, like that of 1997/8, can have a dramatic effect on rainfall amounts. This was shown with particular clarity in the context of Peru (Bendix *et al.* 2000), where normally dry locations suffered huge storms. At Paita (mean annual rainfall 15 mm) there were 1,845 mm of rainfall while at Chulucanas (mean annual rainfall 310 mm) there were 3,803 mm. Major floods resulted (Magilligan and Goldstein 2001).

The Holocene history of El Niño has been a matter of some controversy (Wells and Noller 1999) but Grosjean *et al.* (1997) have discovered more than thirty debris flow events caused by heavy rainfall events between 6.1 and 3.10 Kyr BP in the northern Atacama Desert. The stratigraphy of debris flows has also been examined by Rodbell *et al.* (1999), who have been able to reconstruct their activity over the last 15 Kyr. Between 15 and 7 Kyr BP, the periodicity of deposition was equal to or greater than 15 years and then progressively increased to a periodicity of 2 to 8.5 years. The modern periodicity of El Niño may have been established about 5 Kyr BP, possibly in response to orbitally driven changes in solar radiation (Liu *et al.* 2000). Going back still further, studies of the geochemistry of dated *Porites* corals from the last interglacial of Indonesia have shown that at that time there was an ENSO signal with frequencies nearly identical to the instrumental record from 1856–1976 (Hughen *et al.* 1999).

El Niño events have considerable geomorphological significance. For example, the changes in temperatures of sea water between El Niño and La Niña years have a clear significance for coral reefs (Spencer *et al.* 2000). In 1998, sea surface temperatures in the tropical Indian Ocean were as much as 3–5 °C above normal, and this led to up to 90 per cent coral mortality in shallow areas (Reaser *et al.* 2000). Although other factors may be implicated in coral mortality (e.g. eutrophication, disease, heavy fishing, etc.) large changes in the health of reefs have been noted from remote islands and reefs with low levels of human influence. It would seem that warm conditions between 25 and 29 °C are good for coral growth, but that temperatures above 30 °C are deleterious (McClanahan 2000) and lead to such phenomena as coral bleaching. Bleaching tends to be greatest at shallow depths but a feature of the 1998 event was that it not only caused bleaching of rapidly growing species, but also affected massive species. It also reached to depths as great as 50 m in the Maldives. If global sea temperatures rise as a result of global warming and become closer to the thermal tolerance level of 30 °C, El Niño events of smaller magnitude will be sufficient to cause bleaching. Moreover, the closer the mean sea temperature is to this thermal limit, the longer will be the period for which the tolerances of corals will be exceeded during any El Niño, thereby increasing the likelihood of coral mortality (Souter *et al.* 2000).

Lake levels also respond to El Niño events. El Niño warming in 1997 led to increased rainfall over East Africa that caused Lake Victoria to rise by 1.7 m and Lake Turkana by *c.*2 m (Birkett *et al.* 1999). The abrupt rise in the level of the Caspian Sea (2.5 m between 1978 and 1995) has also been attributed to ENSO phenomena (Arpe *et al.* 2000). Similarly, the 3.7-m rise in the level of the Great Salt Lake (Utah, USA) between 1982 and 1986 was at least partly related to the record rainfall and snowfall in its catchment during the 1982/3 El Niño (Arnow and Stephens 1990). The enormous changes that occur in the areal extent of Lake Eyre in Australia result from ENSO-related changes in inflow, with the greatest flooding occurring during La Niña phases (Kotwicki and Allan 1998).

Some glacier fluctuations are controlled in part by El Niño. Glacier retreat in the tropical Andes can be attributed to increased ablation during the warm phases of ENSO (Francou *et al.* 2000). Conversely, further south, in the southern Andean Patagonia of Argentina, El Niño events have led to increased snow accumulation, causing glaciers to advance so that they create barriers across drainage, creating glacier-dammed lakes (Depetris and Pasquini 2000).

El Niño impacts upon tropical cyclone activity, and the differences in cyclone frequency between El Niño and La Niña years is considerable (Bove *et al.* 1998). For example, the probabilities of at least two hurricanes striking the US is 28 per cent during El Niño years, 48 per cent during neutral years and 66 per cent during La Niña years. There can be very large differences in hurricane landfalls from decade to decade. In Florida, over the period 1851–1996, the number of hurricane landfalls ranged from 3 per decade (1860s, and 1980s) to 17 per decade (1940s) (Elsner and Kara 1999). Given the importance of hurricanes for slope, channel and coastal processes, changes of this type of magnitude have considerable geomorphological significance. Mangroves, for example, are highly susceptible to hurricanes, being damaged by high winds and surges (Doyle and Girod 1997).

Streamflows and sediment yields may also be affected by El Niño events. One area where there have been many investigations of the links between ENSO and streamflow is in the western United States. There is a tendency for the south-west to be wet and the north-east to be dry during the El Niño warm phases (Negative Southern Oscillation Index), and vice versa for La Niña (Cayan *et al.* 1999). There is some evidence that the effect on streamflow is amplified over that on precipitation. A study of sediment yields in southern California showed that during strong El Niño years severe storms and extensive runoff occurred, producing sediment fluxes that exceeded those of dry years by a factor of about five. The abrupt transition from a dry climate to a wet climate in 1969 brought a suspended sediment flux in the rivers of the Transverse Range of 100 million tons, an amount greater than their total flux during the preceding 25-year period (Inman and Jenkins 1999). The wet period from 1978–1983 caused a significant response on alluvial fans and in channels in desert piedmont areas (Kochel *et al.* 1997).

Phases of high sediment yield may themselves have geomorphological consequences. It has been argued, for example, that Holocene beach ridge sequences along the north coast of Peru may record El Niño events that have occurred over the last few thousands of years. The argument (Ortlieb and Machare 1993) is that heavy rainfall causes exceptional runoff and sediment supply to coastal rivers. This, combined with rough sea conditions and elevated sea levels, is potentially favourable for the formation of beach ridge sequences. The high sea levels caused by El Niño, often amounting to 20–30 cm, can contribute to washover of coastal barriers (Morton *et al.* 2000).

Heavy rainfall events associated with ENSO phenomena can cause slope instability. Some of the most distinctive landslides in the south-west of the USA have occurred during El Niño events, and they can be especially serious if the heavy rainfall events occur on slopes that have been subjected to fires associated with previous drought episodes (Swetnam and Betancourt 1990).

On the other hand, exceedingly wet years can in due course cause a great increase in vegetation cover on slopes that may persist for some years and so lead to more stable conditions. In the arid islands of the Gulf of California, for example, plant cover ranges from 0–5 per cent during 'normal' years, but during rainy El Niño periods it rises to 54–89 per cent of the surface available for growth (Holmgren *et al.* 2001). Wet ENSO events can provide rare windows of opportunity for the recruitment of trees and shrubs. Such woodland can be resilient and, once established, can persist.

ENSO can be associated with intensified drought conditions and so can influence the activity of aeolian processes, particularly in areas which are at a threshold for dust entrainment or dune activation. Such areas will be those where in wet years there is just enough vegetation to stabilize ground surfaces. In the USA, dust emissions were greatly reduced in the period 1983–84 following the heavy rainfall of the 1982 El Niño (Lancaster 1997). Likewise, Forman *et al.* (2001) have reconstructed the history of dune movements in the Holocene in the USA Great Plains. They have found that phases of dune activity have been associated with a La Niña-dominated climate state and weakened cyclogenesis over central North America.

References

Arnow, T. and Stephens, D. (1990) Hydrologic characteristics of the Great Salt Lake, Utah, 1847–1986, *US Geological Survey Water-Supply Paper* 2,332, 32.

Arpe, K., Bengtsson, L., Golitsyn, G.S., Mokhov, I.I., Semenov, A. and Sporyshev, P.V. (2000) Connection between Caspian sea level variability and ENSO, *Geophysical Research Letters* 27, 2,693–2,696.

Bendix, J., Bendix, A. and Richter, M. (2000) El Niño 1997/1998 in Nordperu: Anzeichen eines Ökosystem – Wandels? *Petermanns Geographische Mitteilungen* 144, 20–31.

Birkett, C., Murtugudde, R. and Allan, J.A. (1999) Indian ocean climate event brings floods to East Africa's lakes and the Sudd Marsh, *Geophysical Research Letters* 26, 1,031–1,034.
Bove, M.C., Elsner, J.B., Landsea, C.W., Niu, X. and O'Brien, J.J. (1998) Effect of El Niño on US landfalling hurricanes, revisited, *Bulletin American Meteorological Society* 79, 2,477–2,482.
Cayan, D.R., Redmond, K.T. and Riddle, L.G. (1999) ENSO and hydrological extremes in the Western United States, *Journal of Climate* 12, 2,881–2,893.
Depetris, P.J. and Pasquini, A.I. (2000) The hydrological signal of the Perito Moreno Glacier damming of Lake Argentino (Southern Andean Patagonia): the connection to climate anomalies, *Global and Planetary Change* 26, 367–374.
Doyle, T.W. and Girod, G.F. (1997) The frequency and intensity of Atlantic hurricanes and their influence on the structure of the South Florida mangrove communities, in H.F. Diaz and R.S. Pulwarty (eds) *Hurricanes*, 109–120, Berlin: Springer.
Elsner, J.B. and Kara, A.B. (1999) *Hurricanes of the North Atlantic*, New York: Oxford University Press.
Forman, S.L., Oglesby, R. and Webb, R.S. (2001) Temporal and spatial patterns of Holocene dune activity on the Great Plains of North America: megadroughts and climate links, *Global and Planetary Change* 29, 1–29.
Francou, B., Ramirez, E., Cáceres, B. and Mendoza, J. (2000) Glacier evolution in the tropical Andes during the last decades of the 20th century: Chalcaltaya, Bolivia and Antizana, Ecuador, *Ambio* 29, 416–422.
Grosjean, M., Núñez, L., Castajena, I. and Messerli, B. (1997) Mid-Holocene climate and culture changes in the Atacama Desert, Northern Chile, *Quaternary Research* 48, 239–246.
Holmgren, M., Scheffer, M., Ezcurra, E., Gutierrez, J.R. and Mohren, G.M.J. (2001) El Niño effects on the dynamics of terrestrial ecosystems, *Trends in Ecology and Evolution* 16, 89–94.
Hughen, K.A., Schrag, D.P., Jacobsen, S.B. and Hantoro, W. (1999) El Niño during the last interglacial period recorded by a fossil coral from Indonesia, *Geophysical Research Letters* 26, 3,129–3,132.
Inman, D.L. and Jenkins, S.A. (1999) Climate change and the episodicity of sediment flux of small California rivers, *Journal of Geology* 107, 251–270.
Kochel, R.C., Miller, J.R and Ritter, D.F. (1997) Geomorphic response to minor cyclic climate changes, San Diego County, California, *Geomorphology* 19, 277–302.
Kotwicki, V. and Allen, R. (1998) La Niña de Australia – contemporary and palaeo-hydrology of Lake Eyre, *Palaeogeography, Palaeoclimatology, Palaeoecology* 84, 87–98.
Lancaster, N. (1997) Response of eolian geomorphic systems to minor climate change: examples from the southern Californian deserts, *Geomorphology* 19, 333–347.
Liu, Z., Kutzbach, J. and Wu, L. (2000) Modelling climate shift of El Niño variability in the Holocene, *Geophysical Research Letters* 27, 2,265–2,268.
McClanahan, T.R. (2000) Bleaching damage and recovery potential of Maldivian Coral Reefs, *Marine Pollution Bulletin* 40, 587–597.
Magilligan, F.J. and Goldstein, P.S. (2001) El Niño floods and culture change: a late Holocene flood history for the Rio Moquegua, Southern Peru, *Geology* 29, 431–434.
Morton, R.A., Gonzalez, J.L., Lopez, G.I. and Correa, I.D. (2000) Frequent non-storm washover of barrier islands, Pacific coast of Colombia, *Journal of Coastal Research* 16, 82–87.
Ortlieb, L. and Machare, J. (1993) Former El Niño events: records from western South America, *Global and Planetary Change* 7, 181–202.
Reaser, J.K., Pomerance, R. and Thomas, P.P. (2000) Coral bleaching and global climate change: scientific findings and policy recommendations, *Conservation Biology* 14, 1,500–1,511.
Rodbell, D.T., Seltzer, G.O., Anderson, D.M., Abbott, M.B., Enfield, D.B. and Newman, J.H. (1999) An ~15,000-year record of El Niño-driven alluviation in southwestern Ecuador, *Science* 283, 516–520.
Souter, D.W. and Linden, O. (2000) The health and future of coral reef systems, *Ocean and Coastal Management* 43, 657–688.
Spencer, T., Teleki, K.A., Bradshaw, C. and Spalding, M.D. (2000) Coral bleaching in the Southern Seychelles during the 1997–1998 Indian Ocean warm event, *Marine Pollution Bulletin* 40, 569–586.
Swetnam, T.W. and Betancourt, J.L. (1990) Fire-Southern Oscillation Relations in the Southwestern United States, *Science* 249, 1,017–1,020.
Wells, L.E. and Noller, J.S. (1999) Holocene evolution of the physical landscape and human settlement in northern coastal Peru, *Geoarchaeology* 14, 755–789.

A.S. GOUDIE

ELUVIUM AND ELUVIATION

For soils which exist in areas where the water balance is such that rainfall is greater than evaporation, the excess water drains downwards under the influence of gravity. This percolating water can carry material in solution, a process known as leaching. Additionally, material in the form of very fine particles can be moved down in suspension and this is referred to as eluviation – or a 'washing' of particles out of an upper soil horizon, the material being referred to as eluvium. Eluviation can be referred to as mechanical eluviation to distinguish it from losses occuring in solution. When the material becomes redeposited further down the soil profile it is referred to as illuvium, or that material which is washed in to a new lower location by illuviation. The process overall results in the upper soil horizons having a coarser texture, and therefore a greater porosity and permeability, and can result in a finer textured, and sometimes compacted, layer below.

In the lower soil profile, the redeposition of the eluvial material takes place within voids or on the walls of channels, forming a coating of clay round coarser particles in a skin of material where the clay particles are often oriented parallel to each other round the large particle. Such a clay skin is referred to as a cutan, derived from the Latin *cutis*, meaning a skin, coating or rind (cf. cuticle) (Brewer 1964). Such a deposition contributes to the decease of soil pore size and can thus impede further drainage.

Alternatively the eluvial material can be washed out of the soil profile in downslope moving waters such as throughflow or return overland flow. Here the eluvium may be redeposited in or on the soil at the slope foot or washed out of the hillslope system, contributing to the suspended sediment load of rivers and thus forming a constituent of the denudation system of the hillslope in a drainage basin. Whether the eluvial material is redeposited within the soil or reaches the river is largely a matter of the porosity and permeability of the soil and the overall water balance, thus eluviation and hillslope loss to a river is more characteristic of permeable soils in climates with a moderate or high rainfall whereas in less permeable soils and/or with climates with lower rainfall the eluvium is more likely to stay in the soils as redeposited illuvium. Thus, eluviation can form a significant denudation process where there are permeable soils and regoliths. Ruxton (1958) calculated that denudation of hillslopes in the Sudan by eluviation was almost as significant as that by the removal of material in solution, with around 25 per cent of removal by the former and around 35 per cent by the latter.

At the intermediate stage of deposition between the lower soil profile and loss to a river, the formation of clay plains at the foot of slopes along seepage lines can be quite significant. Ruxton (1958) reported such lateral sediment transport and deposition in the form of surface deposits of eluviated clay near the edges of weathered granite domes. The deposits were up to 500 m long, curved round the base of the slope, and up to 150 m wide, though with some longer tongues of deposition where there was evidence of greater water flow and some channelization. Steep (20°) slopes give way sharply to low-angle slopes of deposited fine material below. Here, there was evidently enough rainfall to wash the clay from the higher areas through the bedrock but insufficient runoff to transport the clay further than the slope foot.

Where there are landforms constituted from loose, unstable material, such as sand dunes or even under periglacial conditions with repeated frost heave and downslope movement, eluviation is not a dominant feature as the material is frequently in motion. However, if the material stabilizes – by vegetation growth on dunes or amelioration of climate – then eluviation can occur and this can be referred to as an eluvial phase of development for the soils and associated landform.

References

Brewer, R. (1964) *Fabric and Mineral Analysis of Soils*, Chichester: Wiley.
Ruxton, B.P. (1958) Weathering and sub-surface erosion in granite at the piedmont angle, Balos, Sudan, *Geological Magazine* 95, 353–377.

STEVE TRUDGILL

ENDOKARST

Endokarst consists of the main part of karstic relief and contains carbonate rocks and cave systems (shafts, cavern, etc.). It involves all the underground features of the input karst and of the output karst. It is situated below the EPIKARSTic zone and is fully developed when the karstification is mature (Ford and Williams 1989). It can develop when the acid water from the surface (humic acid from biological activity and carbonic acid from CO_2 exchanges between atmosphere and rainwater) can reach the deepest part of the carbonate (or other karstifiable rocks) layers. It means that the openings of the stratification joints and fractures are large enough to allow the flow of water and suspended material. The development of endokarst is ruled by the competition between dissolution, carbonate precipitation and clogging with non-dissolved particulate material (Rodet *et al.* 1995). Infillings of endokarst conduits are frequently used to date the genesis and the evolution of cave systems (Maire 1990).

The thickness and the lateral extension of the karstifiable layers mark the boundaries of each endokarst. The world's largest cave systems explored by humans are known to exceed 10,000 km^2 and to reach some depths of more than 2,000 m.

References

Ford, D. and Williams, P. (1989) *Karst Geomorphology and Hydrology*, London: Unwin Hyman.
Maire, R. (1990) La haute montagne calcaire, *Karstologia* 3, 731.
Rodet, J., Meyer, R., Dupont J.P., Sayaret, D., Tomat, A. and Viard, J.P. (1995) Relations entre la dissolution des carbonates et le remplissage terrigène dans le karst de la craie en Normandie (France), *Comptes Rendus Academie de Science de Paris*, IIa, 321, 1,155–1,162.

SEE ALSO: chemical weathering; epikarst; ground water; karst; palaeokarst and relict karst

MICHEL LACROIX

ENGINEERING GEOMORPHOLOGY

Engineering geomorphology deals with the geomorphic features of the Earth's surface, with special reference to their engineering properties. These properties include topography, rock units (lithology, rock mass strength, joint spacing, point load range, plasticity characteristics, compaction bearing strength, texture, etc.) soil units, water retention capacity, weathering, mass movement, erosion, etc. The results are very significant in sustainable land management (SLM) through combating processes of land degradation and DESERTIFICATION, and for obtaining a better planning system (see ENVIRONMENTAL GEOMORPHOLOGY).

Engineering geomorphological maps are prepared in a composite form from which collected, processed and stored data, required for a particular project, are extracted and analysed. Derivative maps are also obtained for specific purposes. An engineering geomorphological map (EGM) depicts the morphological and engineering properties of the terrain. It is very useful during the phase of policy or project formulation and implementation, the phase of development and construction, and the continuing management of a development, particularly in the context of civil engineering (see GEOMORPHOLOGICAL MAPPING; TERRAIN EVALUATION).

Various studies have been made in the field of APPLIED GEOMORPHOLOGY that particularly emphasize engineering applications (Cooke and Doornkamp 1974; Hails 1977; Verstappen 1982; Jha and Mandal 1997).

The methodology for the preparation of an engineering map (Table 16) involves three phases: (1) pre-fieldwork; (2) fieldwork; and (3) post-fieldwork. The pre-field phase includes the preparation of the base map from topographical sheets and cadastral maps. Delineation of the area is also done with high-resolution satellite imagery and aerial photographs at a 1:5,000 to 1:25,000 scale.

Table 16 Phases in engineering geomorphological mapping

Pre-fieldwork
Objectives
Framing of delineation rule
Delineation of the study area (topographical and cadastral maps, aerial photographs, and satellite imagery)
Classification criteria
Selection of engineering geomorphological properties
Selection of sites for collection of rock samples
Fieldwork
Field checks, scanning of engineering geomorphological properties
Collection of rock samples, soils, weathering, mass movement and erosional characteristics of rills and gullies
Rock fall, landslides, slope failures and scree zones
Post-fieldwork
Discussion
(Addition/alteration, etc.)
Field mapping corroboration

In the fieldwork phase all the rock units, soil units, tectonic elements, and active geomorphic processes are investigated. Also noted is the water retention capacity of the mass. Ultimately, the areal coverage and locations of occurrences are marked on the base map. The topographic features marked from the toposheets are also updated during field investigation. Rock samples are collected from different litho-stratigraphic units for laboratory analysis to obtain their geo-engineering properties.

The post-fieldwork phase involves laboratory testing of rock samples, transfer of field data relating to mass movement, weathering and erosion, and preparation of the final map and its interpretation. The final map can be prepared by transferring and plotting data obtained from different sources in a synthesized manner, and by using different symbols and colours. During this phase of work, aerial photographs and satellite images are consulted, in addition to the fieldwork, to identify and demarcate the exact location and boundaries

of different rock mass strength units, soil units, zones of weathering and areas of active mass movements and erosion, etc. Finally, the engineering geomorphological map for the study area can be prepared.

Engineering geomorphological map preparation (see Figure 54) includes the following physical parameters: topography; rock units; tectonics; soil units; weathering; mass movement; and erosion. The engineering geomorphological map gives special emphasis to the following aspects of the rock mass: strength of intact rock; joint spacing; width of joints; bedding planes; gauge or infilling; the materials and water movements within the rock; point load range; water retention capacity; and compaction bearing strength.

Considering the above attributes of the rock units, the study area can be divided into four categories by superimposing the layering of information on the base map using a Geographic Information System (GIS).

1 Low rock mass strength unit (Rlo)
2 Medium rock mass strength unit (Rme)
3 High rock mass strength unit (Rhi)
4 Very high rock mass strength unit (Rvhi).

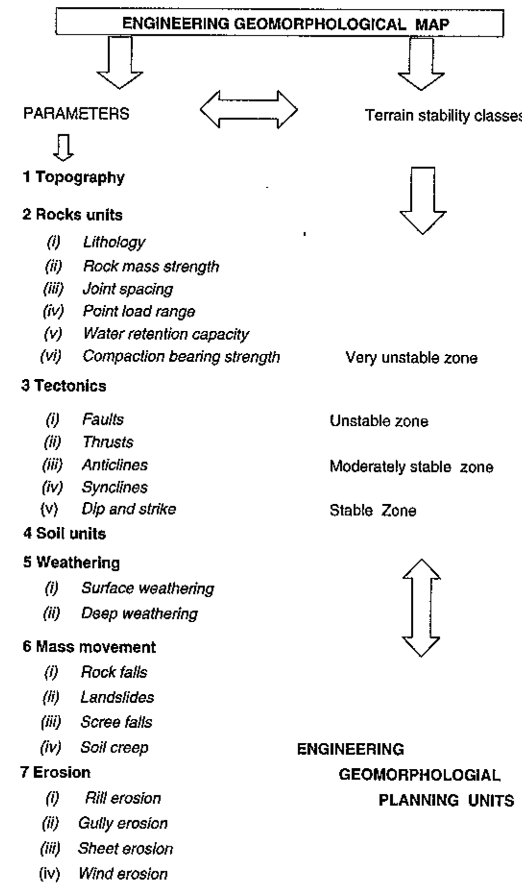


Figure 54 Parameters of the engineering geomorphological map

The engineering geomorphological map shows that all the parameters are functionally interrelated in the sample area. The result can be presented in four terrain stability classes: very unstable zone; unstable zone; moderately stable zone; and stable zone.

An engineering geomorphological map is an important tool in planning and development. It indicates the stability of terrain which is very significant in civil constructions, agriculture, industry, transportation networks and settlement establishment. As this type of map records all sorts of topographical, morphological, and geo-engineering data, so the development and planning of an area should be based on this type of mapping in which the land use and other planning aspects can be regulated according to the stability/suitability of the terrain.

References

Cooke, R.U. and Doornkamp, J. (1974) *Geomorphology in Environmental Management: An Introduction*, Oxford: Clarendon Press.

Hails, J.R. (1977) *Applied Geomorphology*, Amsterdam: Elsevier.

Jha, V.C. and Mandal, U.K. (1997) Drainage Basin in Environmental Management: A Case Study of Eastern Nayer Basin, U.P. Himalayas, in P. Nag, V.K. Kumra and J. Singh (eds) *Geography and Environment*, Vol. 2, 204–225, New Delhi: Concept.

Verstappen, H.Th. (1982) *Applied Geomorphology*, Amsterdam: Elsevier.

VIBHASH C. JHA

ENVIRONMENTAL GEOMORPHOLOGY

Environmental geomorphology is the practical use of geomorphology for the solution of problems where humans wish to transform or to use and change surface processes. According to Coates (1971, 1972–1975), this discipline involves the following issues and themes:

1. the study of geomorphic processes and terrain that affect man, including hazard phenomena such as floods and landslides;
2. the analysis of problems where man plans to disturb or has already degraded the land–water ecosystem;
3. human utilization of geomorphic agents or products as resources, such as water or sand and gravel;
4. how the science of geomorphology can be used in environmental planning and management.

Many other researchers have dealt with environmental geomorphology, in the discussion of both specific topics and the various applications of geomorphology in the forms of APPLIED GEOMORPHOLOGY, ENGINEERING GEOMORPHOLOGY and also engineering geology. It is not necessary to review the available literature here; it is sufficient to mention Tricart (1962, 1973, 1978), Verstappen (1968, 1983), Craig and Craft (1982) and Cendrero *et al.* (1992), among others.

More recently, Panizza (1996) defined *environmental geomorphology* as that area of Earth sciences which examines the *relationships between man and environment*, the latter being considered from the *geomorphological* point of view. It should be further specified that *environment*, in general, is defined (Panizza 1988) as the 'range of physical and biological components that have an effect on life and on the development and activities of living organisms'.

The geomorphological components of the environment may be schematically subdivided into: *geomorphological resources*; and *geomorphological hazards*. Geomorphological resources include both raw materials (related to geomorphological processes) and landforms: both of which are useful to man or may become useful depending on economic, social and technological circumstances. For instance, littoral deposits can become important, economically valuable and considered as geomorphological resources when used for sand quarrying. Similarly, a sea beach can acquire value and be considered as a geomorphological resource when utilized as a seaside resort. A landform can be considered a resource also from the scientific and cultural viewpoint: for example, a marine cliff can be seen as a model of geomorphological evolution.

GEOMORPHOLOGICAL HAZARDS can be defined (Coates 1972–1975) as the 'probability that a certain phenomenon of geomorphological instability and of a given magnitude may occur in a certain territory in a given period of time'. For example, in any one area, the possibility of a certain landslide occurring over a 50-year time span may be assessed. Hazard is therefore a function of the intensity/magnitude and of the frequency/probability of the phenomenon (Varnes 1984). The term 'susceptibility' as used in many mapping procedures (e.g. Brabb *et al.* 1972) corresponds to hazard by equating spatial probabilities to temporal probabilities.

In the context of the relationships with the environment, man represents: ***human activity***; and ***area vulnerability***. Human activity is the specific action of man which may be summarized under the headings of hunting, grazing, farming, deforestation, utilization of natural resources, engineering works, etc. Man's interventions take place essentially on that thin layer of the Earth's surface which makes up the interface between atmosphere and lithosphere where most energy exchanges and complex phenomena take place (Piacente 1996). Hardly ever are these phenomena confinable within preconstituted and rigid schemes, but nevertheless they can be summarized as follows (Castiglioni 1979): artificial forms, directly modelled by man's activities; works aiming to divert, correct or upgrade natural processes; modifications of natural phenomena, indirectly resulting from man's activities.

Area vulnerability is the complex of the inhabitants and all things that exist as a result of the work of man in a given area and which may be directly or indirectly sensitive to material damage. Included in this complex, we find the population, buildings and structures, infrastructures, economic activity, social organization and any expansion and development programmes planned for an area. In short, it corresponds to an 'exposed element'. Vulnerability can also be defined (Varnes 1984; Einstein 1988) as the level of potential damage (ranging from 0 to 1) to a given exposed element, which is subject to a possible or real phenomenon of a given intensity: we prefer the first definition, which does not imply elements already included in the definition of hazard.

Considering the relationships between geomorphological environment and man, two main possibilities can be examined (Panizza 1992) (Figure 55):

1. Geomorphological resources in relation to human activity, where geomorphological environment is regarded as mainly passive in relation to man (active); in other words, a resource may be altered or destroyed by human activity (e.g. a mountain landscape that has been modified by a bulldozer). We define as *impact* these consequences of human activity on geomorphological resources. It consists of the physical, biological and social changes that human intervention brings about in the environment, the latter term being intended in its geomorphological elements. Therefore, this impact equals the 'product' of human activity and geomorphological resources.
2. Geomorphological hazard in relation to area vulnerability, where geomorphological environment is regarded as mainly active in relation to man (passive): in other words, a hazard may alter or destroy some buildings or infrastructures (e.g. a landslide or river

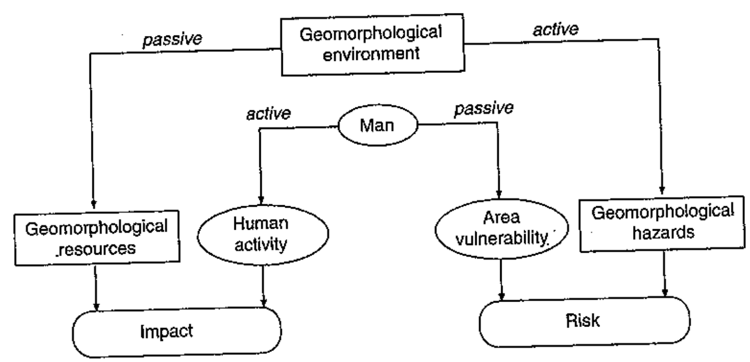


Figure 55 Relationships between geomorphological environment and man

erosion that cause road collapse). We define as *risk* these consequences of geomorphological hazard on a situation of area vulnerability. It is a natural risk connected to a geomorphological hazard: the term refers to the probability that the economic and social consequences of a particular phenomenon reflecting geomorphological instability will exceed a certain threshold. Therefore, this risk is equal to the 'product' of geomorphological hazard and an area's social and economic vulnerability. It corresponds to the term 'specific risk' by Varnes (1984), which expresses the loss due to a particular natural process.

References

Brabb, E.E., Pampeyan, E.H. and Bonilla, M.G. (1972) *Landslide susceptibility in San Mateo county, California*, US Geological Survey Miscellaneous Field Studies Map MF-360, scale 1:62,500.
Castiglioni, G.B. (1979) *Geomorfologia*, Torino: UTET.
Cendrero, A., Luttig, G. and Wolff, F.C. (eds) (1992) *Planning the Use of Earth's Surface*, Berlin: Springer Verlag.
Coates, D. (ed.) (1971) *Environmental Geomorphology*, Proceedings Symposium State University of New York, Binghamton.
——(ed.) (1972–1975) *Environmental Geomorphology and Landscape Conservation*, 3 vols, Stroudsburg, PA: Hutchinson and Ross.
Craig, R.G. and Craft, J.L. (eds) (1982) *Applied Geomorphology*, London: Allen and Unwin.
Einstein, H.H. (1988) *Landslide Risk Assessment Procedure*, International Symposium on Landslides, Lausanne 2, 1,075–1,090.
Panizza, M. (1988) *Geomorfologia applicata. Metodi di applicazione alla Pianificazione territoriale e alla Valutazione d'Impatto Ambientale*, La Nuova Roma: Italia Scientifica.
——(1992) *Geomorfologia*, Bologna: Pitagora.
——(1996) *Environmental Geomorphology*, Amsterdam: Elsevier.
Piacente, S. (1996) Man as geomorphological agent, in M. Panizza (ed.) *Environmental Geomorphology*, Amsterdam: Elsevier.
Tricart, J. (1962) *L'Epiderme de la Terre. Esquisse d'une géomorphologie appliquée*, Paris: Masson.
——(1973) La géomorphologie dans les études intégrées d'aménagement du milieu naturel, *Annales de Géographie* 82, 421–453.
——(1978) *Géomorphologie applicable*, Paris: Masson.
Varnes, D.J. (1984) *Landslide Hazard Zonation: A Review of Principles and Practice*, Paris: Unesco.
Verstappen, H.Th. (1968) Geomorphology and Environment, inaugural address, Delft: Waltman, 1–23.
——(1983) *Applied Geomorphology*, Amsterdam: Elsevier.

MARIO PANIZZA

EPEIROGENY

In his monograph on Lake Bonneville, Gilbert (1890: 340) formalized the definition of certain tectonic terms: 'Displacements of the Earth's crust which produce mountain ridges are called orogenic... the process of mountain formation is orogeny, the process of continent formation is epeirogeny, and the two collectively are diastrophism.' Bloom (1998: 43) concurred that there was need to describe non-orogenic tectonism (i.e. tectonic movements not associated with mountain belts) and so redefined epeirogeny as: 'continental vertical tectonic movement of low amplitude relative to its wavelength, not within an orogenic belt, that does not deform rocks or the land surface to an extent that is measurable within a single exposure.'

Such broad movements can be either positive (uplift) or negative (subsidence). Epeirogenic uplift can be attributed to MANTLE PLUMES beneath broad areas of continental crust and to such processes as glacio-isostasy. Rates of epeirogeny have generally been thought to be one or two orders of magnitude lower over similar time intervals to rates of orogeny, but studies of present rates of neotectonism suggest this is not invariably the case. In areas of active epeirogeny, river incision may occur (Wisniewski and Pazzaglia 2002).

References

Bloom, A.L. (1998) *Geomorphology*, 3rd edition, New Jersey: Prentice Hall.
Gilbert, G.K. (1890) Lake Bonneville, *United States Geological Survey Monograph* No. 1.
Wisniewski, P.A. and Pazzaglia, F.J. (2002) Epeirogenic controls on Canadian River incision and landscape evolution, Great Plains of Northeastern New Mexico, *Journal of Geology* 110, 437–456.

A.S. GOUDIE

EPIKARST

The epikarst is the uppermost highly weathered layer of karst bedrock beneath the soil (Klimchouk 2000). It is also known as the subcutaneous zone (Williams 1983). Where there is no soil there is still an epikarst, for example beneath limestone pavements and alpine karrenfeld. The epikarst develops because rainwater is acidified by dissolving carbon dioxide in the atmosphere and especially in the soil, thereby producing weak carbonic acid. On percolating downwards from the surface into the bedrock, this water accomplishes most of its dissolutional work within 10 m of the surface, i.e. close to its main source of carbon dioxide. The result is that fissures in the limestone are especially enlarged by corrosion near the surface but taper with depth. Consequently, infiltration of rainwater into the karst is initially rapid, but vertical water flow encounters increasing resistance with depth as fissures become narrower and less frequent. This produces a bottleneck effect after particularly heavy rain, resulting in temporary storage of percolation water in a perched epikarstic aquifer.

Joints, faults and bedding-planes vary spatially within the rock because of tectonic history and variations in lithology. As a result the frequency and interconnectedness of fissures available to transmit flow also varies. Nevertheless, near the surface there is considerable interconnectedness in the horizontal plane; so recharging rainwater tends to be homogenized by lateral mixing. However, in the vertical plane some fissures are more favourable for vertical percolation than others, for example master joints that penetrate numerous beds and especially where several joints intersect. As a result these fissures develop as principal drainage paths. Water in the epikarstic aquifer flows laterally towards them and, as a result, they are subjected to still more dissolution by a positive feedback mechanism and so vertical permeability is enhanced. Water captured within the zone of influence of particular drainage routes becomes increasingly isolated as it percolates downwards from water elsewhere in the epikarst, and so despite the early homogenization it gradually acquires a water quality that reflects the residence time in the epikarst.

The saturated zone in the epikarst is especially well developed after heavy rain, when the epikarstic saturated zone is suspended like a perched aquifer above the main phreatic zone in the karst. The piezometric surface (water table) of the epikarst draws down over a preferred leakage path similar to the cone of depression in the water table over a pumped well. Streamlines adjust and resulting flow within the epikarst is centripetal and convergent on the drainage zone. The diameter of any solution doline that ultimately develops as a consequence of the focused dissolution is determined by the radius of the epikarstic drawdown cone.

References

Klimchouk, A. (2000) The formation of epikarst and its role in vadose speleogenesis, in A.B. Klimchouk, D.C. Ford, A.N. Palmer and W. Dreybrodt (eds) *Speleogenesis: Evolution of Karst Aquifers*, Huntsville: National Speleological Society, 91–99.
Williams, P.W. (1983) The role of the subcutaneous zone in karst hydrology, *Journal of Hydrology* (Netherlands) 61, 45–67.

PAUL W. WILLIAMS

EQUIFINALITY

Equifinality is the principle which states that morphology alone cannot be used to reconstruct the mode of origin of a landform on the grounds that identical landforms can be produced by a number of alternative processes, process assemblages or process histories. Different processes may lead to an apparent similarity in the forms produced. For example, sea-level change, tectonic uplift, climatic change, change in source of sediment or water or change in storage may all lead to river incision and a convergence of form. The usage of the term in this way stems from Chorley (1962) but the related concept of converging landforms was developed earlier by Mortensen (1948), who pointed out that there are many convergences in the landforms of arid and polar regions even though their climates (and therefore by implication their geomorphic process assemblages) are so different.

Perhaps one of the better illustrations of the problem of equifinality concerns the origin of TORS. Four principal theories are held to explain identical landforms: (1) subaerial weathering causes spheroidal modification to the morphology of outcrops produced by differential erosion of the bedrock; (2) exposure of tors is due to a two-stage process: a period of prolonged subsurface groundwater weathering leading to decay of closely jointed rock and spheroidal modification of larger blocks, followed by a period of erosional stripping leading to exposure of the tor at the surface; (3) reduction in area of larger inselbergs by scarp retreat and the formation of pediments; and (4) tors are isolated as a result of freeze–thaw weathering, followed by solifluction over permafrost in a periglacial climate. In the last analysis, it is probable that all four processes are reasonable alternative explanations of the origin of this landform. It is generally accepted that no final descriptive definition is without ambiguity. Hence it is not possible to argue from the presence of this landform alone that a certain sequence of genetic events has occurred. This is the classical concept of equifinality.

Brunsden (1990) in discussing his Proposition 10 – the ability of a landscape to resist impulses of change tends to increase with time – notes that in spite of the existence of complex causes and complex responses in geomorphology 'there is within any tectono-climatic domain a tendency toward an all pervading unity and a repetitive but characteristic geometrical order and regularity.' He explains this tendency as resulting from preferential selection of stable forms; exponential decrease in rate of change; increasing effectiveness of barriers to change; constancy of process; persistence; convergence; over-relaxed systems; self-propagation; preferential fabric relief patterns; and process smoothing and extreme event accumulation.

Haines-Young and Petch (1983) suggest that the concept has been misused in that geomorphologists have invoked equifinality in order to avoid the hard question of specific mode of origin of the landforms in question and, they claim, too rapid an acceptance of equifinality may inhibit the development of general laws or may lead to detailed differences of form being overlooked. They suggest a redefinition of the concept as follows: 'a single landform type is said to exhibit equifinality when it can be shown to arise from a range of initial conditions through the operation of the same causal processes' (1983: 465). In this context, they commend Culling (1957) for his use of the graded stream as an example of equifinality in the sense that whatever the initial conditions, a graded stream will display a similar long profile.

Culling (1987) suggests that the word 'equifinality' is no longer useful because advances in our understanding of dynamic systems have opened up a new and richer world with its own more flexible vocabulary. The idea that a system will strive to arrive at similar positions in phase or state space despite differing initial conditions has become familiar to students of general systems theory. The existence of strange attractors in non-linear dissipative dynamic systems is also reminiscent of the older idea of equifinality, but there are equal evidences of chaotic motion in systems that are fully determined and predictable with accuracy in the immediate future. Because of the ubiquity of noise, all stable systems are transient.

It is the recognition of complicated periodic behaviour at points far from equilibrium and its interaction with strange attractors that upsets a simple definition of equifinality. Culling proposes a complex topology of degrees of equifinality, which depart from the definition of equifinality *sensu stricto*. That definition remains 'that upon perturbation a system will eventually return to its initial position'. It is important to realize that such a condition is itself transient. Therein lies the essence of the flexibility of the new approach to the concept of equifinality.

Phillips (1995) discusses the value of viewing landscape evolution as an example of self-organization. Self-organization or self-regulation depend upon the dominance of negative feedback in the system. He points out that geomorphic systems give evidence of both self-organization (as in the case of at-a-station hydraulic geometry) and non-self-organization (as in soil landscape evolution). The challenge for geomorphology is how to distinguish between such systems. Self-similarity and equifinality are incompatible with non-self-organizing systems.

Culling (1987) argues for a variety of looser definitions of equifinality than that of the strict definition above. He suggests that a quasi-equifinality can be defined to include approximate return to initial conditions, by placing a restriction on the allowable magnitude of perturbation, by defining a system domain whose manifold has several local minima and in accepting the retention of certain ergodic and topologic properties as adequate criteria for equifinality.

The debate is reminiscent of the dynamic equilibrium debate of thirty years ago, but with two developments: the level of analytical sophistication has increased and, perhaps more importantly, the debate is open ended and does not yield a unique conclusion. 'In looking once again at the concept of equifinality, it is as if we had opened some magic casement to find, between chance and necessity, one dimension and the next, a whole new world of chaotic motions, strange attractors and periodic windows. With a wild surmise we gaze upon an ocean of discovery between two continents previously thought contiguous' (Culling 1987: 69).

References

Brunsden, D. (1990) Tablets of stone: toward the ten commandments of geomorphology' *Zeitschrift für Geomorphologie Supplementband* 79, 1–37.

Chorley, R.J. (1962) Geomorphology and general systems theory, *US Geological Survey Professional Paper* 500-B.

Culling, W.E.H. (1957) Multicyclic streams and the equilibrium theory of grade, *Journal of Geology* 65, 259–274.

——(1987) Equifinality: modern approaches to dynamical systems and their potential for geographical thought, *Institute of British Geographers Transactions* N.S. 12, 57–72.

Haines-Young, R.H. and Petch, J.R. (1983) Multiple working hypotheses: equifinality and the study of landforms, *Institute of British Geographers Transactions* N.S. 8, 458–466.

Mortensen, H. (1948) Das morphologische Harteverhaltnis Hornfels-Granit, in *Harz. Nachrichten Akademische Wissenschaften Göttingen, Mathematische-Physische Klass*, 8–20.

Phillips, J.D. (1995) Self-organization and landscape evolution, *Progress in Physical Geography* 19, 309–321.

SEE ALSO: non-linear dynamics; tor

OLAV SLAYMAKER

EQUILIBRIUM LINE OF GLACIERS

The position on a glacier where seasonal accumulation equals seasonal ablation is termed the equilibrium line (refer to Figure 56). At this area on a glacier the net mass balance is at zero (no ice mass is lost or gained at this point). A glacier gains mass during winter as snow falls, causing a positive annual mass balance above the equilibrium line. Below the equilibrium line, the annual mass balance of glaciers is negative due to ablation during the summer melt season (Sugden 1982).

The equilibrium line altitude (ELA) for a particular glacier budget year is considered synonymous with the end of summer snowline (EOSS). The altitude of the annual EOSS averaged over many years, defines the steady-state ELA. The annual snowline position with respect to the long-term or steady-state ELA can be used as a surrogate or index of the annual mass balance changes of a glacier. Changes in glacier mass balance are a direct, undelayed response to changes in atmospheric conditions (Fitzharris *et al.* 1997) and hence can be a useful indicator of larger scale changes in global climate.

A commonly used method to work out the ELA calculates the area of the glacier, comparing this to the accumulation area. This is termed the accumulation area ratio (AAR). Glacier studies worldwide have demonstrated that the AAR for glaciers with stable ELA has a value of around 0.6 (Lowe and Walker 1997). Where it is possible to estimate the extent of late Pleistocene glaciers, this method can describe the change in ELA compared with the present, and consequently the changes in climate over time.

References

Fitzharris, B.B., Chin, T.J. and Lamont, G.N. (1997) Glacier balance fluctuations and atmospheric circulation patterns over the Southern Alps, New Zealand, *International Journal of Climatology*, 17, 745–763.

Lowe, J.J. and Walker, M.J.C. (1997) *Reconstructing Quaternary Environments*, 2nd edition, 43–44, Harlow: Addison-Wesley Longman.

Sugden, D. (1982) *Arctic and Antarctic, A Modern Geographical Synthesis*, New Jersey: Barnes and Noble.

BLAIR FITZHARRIS

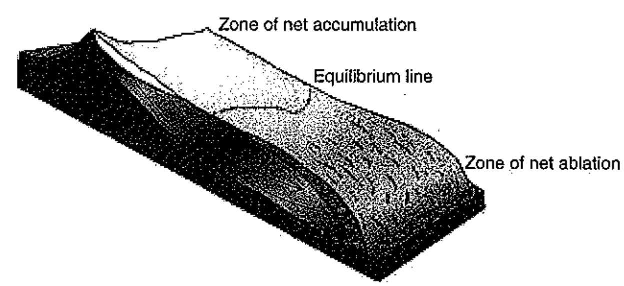


Figure 56 The longitudinal profile of a valley glacier, showing the area of seasonal accumulation and ablation, and the equilibrium line

EQUILIBRIUM SHORELINE

A beach face is in a state of DYNAMIC EQUILIBRIUM when the same amount of sediment is moved landwards by the stronger uprush as is moved seawards by the weaker backrush. This is accomplished by adjustments to the gradient and shape of the BEACH. The gradient is largely determined by the amount of water that percolates into the beach and is lost to the downrush, which is primarily a function of grain size. Pebble or shingle beaches (see SHINGLE COAST) are particularly steep because of rapid percolation and the consequently very weak backrush. Much less water percolates into fine-grained sandy beaches, however, and the weak gravitational effect of a gently sloping beach face is therefore able to compensate for small differences in the onshore and offshore transport rates. There is also a tendency for beach gradient to decrease as wave steepness increases, presumably because the greater velocity of the uprush makes it easier to carry sediment up the slope. The gradient of beaches, or portions of beaches, with the same grain size can therefore vary according to differences in exposure and wave steepness, while temporal variations in wave steepness explain why beach gradient changes during and following storms. There is also evidence to suggest that beach slope is partly determined by the height or energy of the WAVE, which would explain why, for the same wave steepness, slopes are generally greater in low than in high energy environments. The proportion of heavy minerals in a beach, which increases the resistance to removal by backrush, may also be significant in determining the equilibrium gradient.

It is difficult to know if beaches are in quasi-equilibrium, and it can also be difficult to define the slope of beaches that consist of more than one slope element, although it is usually measured in the swash zone where it is essentially linear. A variety of descriptive, empirical and mathematical models are concerned with the relationship between equilibrium beach slope, grain size and wave parameters. The first attempts to model beach gradient were largely statistical and concerned with correlations with wave and sedimentological parameters. Analytical models solve equations for equilibrium gradient on the assumption that there is no net sediment transport, whereas iterative models simulate beach development until an equilibrium slope has developed.

Bruun (see BRUUN RULE) suggested that the shape of equilibrium beach profiles can be represented by the power law:

$$h_x = A\,x^{2/3}$$

where h_x is the depth at a distance x offshore of the mean water line and A is a scale parameter that is largely determined by grain size or fall velocity. If there has been significant sorting, however, coarser sediment can make the shoreward portions of beaches steeper than model predictions, and finer sediment can make the seaward portions more gentle than predicted. Although other workers have provided alternate expressions for the geometry of the equilibrium profile, the use of a single equilibrium equation to represent all beach profiles has been criticized, and the concept of a profile of equilibrium has been questioned.

Beaches adjust their equilibrium morphology and sediments with variations in waves, tides and other influences (Short 1999). Two profiles represent the extremes of a fairly predictable range of forms that can be assumed according to the size or power of the waves. The distinction has been made between reflective profiles with wide berms or swash bars and steep foreshore slopes, sometimes with steps at the breaker line, and dissipative profiles with gentler foreshore slopes and longshore submarine bars (see BAR, COASTAL). Reflective profiles change into dissipative profiles during storms, when large waves move sediment seawards, whereas the reverse occurs when smaller swell waves move sediment back onshore. Frequent changes in wave power are responsible for cycles that are frequently much shorter than that between the two extremes, and wave environments therefore tend to generate globally distinctive beach state characteristics. The mid-latitudes, for example, have persistently high wave power and the beaches are generally kept in a highly dissipative state, whereas beaches in low swell or sheltered environments are normally in reflective states. Beach states also change with tidal level, and the microtidal model has to be modified for areas with a high tidal range. Equilibrium beach profiles and beach states have been modelled as a function of the relative tidal range (RTR) – the ratio of tidal range to breaker height and the dimensionless fall velocity of the sediment, with high RTR values representing tide-dominance and low values wave-dominance.

Coastlines trend towards an equilibrium state in the longshore as well as in the cross-shore direction. The distinction can be made between swash- and drift-aligned equilibrium beach forms (Plates 39 and 40). Swash-aligned beaches are parallel to the incoming wave crests and net LONGSHORE (LITTORAL) DRIFT is at a minimum. Drift-aligned beaches are parallel to the line of maximum drift and sediments can be carried great distances in one direction. Swash-aligned beaches are associated with irregular coasts where longshore transport is impeded, and the important wave trains reach the shoreline almost normally. Drift-aligned beaches develop where the initial coastal outline is fairly regular, or where important sediment-moving waves approach the coast at an angle.

Drift alignment and dynamic beach equilibrium require a constant supply of sediment, as for example when a coastal cell is coupled to the

Plate 39 Swash-aligned beach at San Martinho do Porto, Portugal

Plate 40 Drift-aligned beach, St Petersburg, Florida

mouth of an ESTUARY. Static beach equilibrium can be attained in several ways (Carter 1988):

1. through swash-alignment, when sediment movement is restricted to cross-shore transport;
2. by strong wave height gradients or the interaction of two wave trains causing the longshore current velocity to become zero; and
3. by the alongshore grading of beach sediments in such a way that the strength of the current at each place is too low to entrain it.

All three situations are common, and in some cases equilibrium is attained through a combination of options.

References

Carter, R.W.G. (1988) *Coastal Environments*, London: Academic Press.

Short, A.D. (ed.) (1999) *Handbook of Beach and Shoreface Morphodynamics*, Chichester: Wiley.

Further reading

Trenhaile, A.S. (1997) *Coastal Dynamics and Landforms*, Oxford: Oxford University Press.

ALAN TRENHAILE

EQUILIBRIUM SLOPE

Equilibrium slopes are hillslopes which are characterized by an equilibrium of forces that compensate each other. An equilibrium slope exists 'if the amount of material that is removed from an areal unit of the surface per time unit is equal to the amount of material that is supplied to this areal unit during the same time' (Ahnert 1994). This definition follows the conception that a geomorphic system, e.g. a hillslope, is under equilibrium conditions if the 'mass budget' of that system does not change. Hillslopes are in equilibrium conditions if the processes acting upon the hillslope are in equilibrium. Each change of this equilibrium will result in adjustments of the acting processes towards a new equilibrium.

This definition goes back to the fundamental work of Grove Karl Gilbert on the Henry Mountains (Gilbert 1877). In this publication Gilbert introduced the concept of equality of action: 'Erosion is most rapid where the resistance is least, and hence as the soft rocks are worn away the hard are left prominent. The differentiation continues until an equilibrium is reached through the law of declivities (slope gradient). When the ratio of erosive action as dependent on declivities becomes equal to the ratio of resistances as dependent on rock character, there is equality of action.' This situation is called 'dynamic equilibrium' because the system equilibrium is attained by mechanisms of self-regulation, where a change of process components caused by a change of input will result in a compensation between these process components by negative feedbacks (see DYNAMIC EQUILIBRIUM). The negative feedback between processes governing the system causes adjustments to the changes of inputs (e.g. a climatic change or a BASE LEVEL lowering) towards a new equilibrium. In this way the two central aspects (1) of mass transport rates and (2) negative feedback mechanisms were established in geomorphology.

Gilbert (1877) distinguished two types of transport laws of hillslopes: weathering-limited and transport-limited regolith removal. Weathering-limited transport occurs where the weathering rate is lower than the transport capacity of the hillslope forming processes, so that the regolith is removed and slope development is related to the weathering rate of rocks. In this case the slope system is in non-equilibrium, the inverse of equilibrium. The material supply by different slope processes (weathering, slope wash, soil creep, etc. from upslope) is smaller than the potential rate of removal. These form elements can be found in arid and semi-arid environments, in mountain areas and on free faces of cliffs and all stream channels in bedrock. Transport-limited transport occurs where the weathering rate is higher than the transport capacity of the hillslope-forming processes, so that regolith accumulates and the transport processes operate at their full capacities.

The concept of equilibrium slopes has been applied to numerous investigations in geomorphology related to slope evolution by river incision and undercutting followed by mass movements if internal frictional threshold angles of the regolith are crossed. Examples are given, for instance, by the pioneering research of Strahler (1950), who related statistically maximum valley-side slope gradients with the frictional threshold angles of up to 1-m thick regolith cover. Similar equilibrium approaches were published by Young (1972) and Carson (1975).

Further approaches are related to finding characteristic form slope profiles (equilibrium profiles) for a range of transport processes by empirical modelling transport capacity relationships. Kirkby (1971) used this relationship to derive characteristic equilibrium slope profiles for soil creep, soil wash without and with gullying and rivers. Ahnert (1976) developed a more complex computer model which generates five different equilibrium slope profiles related to splash (convex), suspended load-wash (convex–straight–concave), point-to-point wash (rolling) (convex–straight), plastic flow and viscous flow (convex–straight) processes. These models are based on equilibrium assumptions concerning mass transport rates of different processes and feedback mechanisms between them.

There are extreme events that destroy the equilibrium on hillslopes, e.g. by removing the regolith by extreme rainfall, landsliding, gully erosion or vegetation change. In this situation the process rates change significantly in time, which means that the system is in a state of disequilibrium. If the entire slope system has the tendency towards a dynamic equilibrium, negative feedbacks adjust the slope after these external impacts. The period of recovery from this event depends on the constitution of the system itself and on the magnitude of the external impact. If the regolith coverage has been removed and the bare rocks are exposed, system response will result in an adjustment by an increase of the rate of the weathering processes.

In a recent debate Ahnert (1994) and Thorn and Welford (1994) reviewed different equilibrium definitions and concepts in geomorphology. The authors were especially concerned with a high degree of confusion generated by different types of equilibrium concepts used. Based on the very clearly defined concept of Ahnert (1994), Thorn and Welford (1994) suggested the use in the future of a mass-based equilibria concept which is based on field data, namely mass volume and mass flux. They suggested one should abandon the term 'dynamic equilibrium' and use the term 'mass flux equilibrium' to avoid associations with former definitions and concepts and to reach a clearer coupling with other disciplines.

References

Ahnert, F. (1976) A brief description of a comprehensive three-dimensional process-response model of landform development, *Zeitschrift für Geomorphologie, N.F. Supplementband* 25, 29–49.
——(1994) Equilibrium, scale and inheritance in geomorphology, *Geomorphology* 11, 125–140.
Carson, M.A. (1975) Threshold and characteristic angles of straight slopes, *Proceedings of the 4th Guelph Symposium on Geomorphology*, 19–34, Norwich: Geo Books.
Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, Washington, DC: US Geological and Geographical Survey.
Kirkby, M.J. (1971) Hillslope process-response models based on the continuity equation, *Institute of British Geographers Special Publication* 3, 15–30.
Strahler, A.N. (1950) Equilibrium theory of erosional slopes approached by frequency distribution analysis, Part I and Part II, *American Journal of Science* 248, 673–696, 800–814.
Thorn, C.E. and Welford, M.R. (1994) The equilibrium concept in geomorphology, *Annals of the Association of American Geographers* 84, 666–696.
Young, A. (1972) *Slopes*, London: Longman.

SEE ALSO: dynamic equilibrium; equilibrium shoreline; hillslope, form; hillslope, process

RICHARD DIKAU

ERGODIC HYPOTHESIS

Ergodicity is an idea developed in physics. In studying the movement of molecules in a macroscopic system (such as a room full of air), physicists faced a difficult problem: innumerable molecules move very fast compared with the time taken to observe them. To overcome this problem, they devised the ergodic theorem and the ergodic hypothesis, the word ergodic coming from the Greek *ergon* ('work' or 'energy') and *hodos* ('road') and meaning a 'path of constant energy'.

To appreciate the ideas behind ergodicity, take the case of people in a maze. Now, a maze is a network consisting of a number of links joined at nodes. Imagine one person entering the maze, which has no exit, and wandering around long enough to have entered all possible links at least once. By keeping a record of the path taken by the person, it is possible to calculate two probabilities. First, is the probability of the person's being in a particular link after a given time. Second, is the probability that the person has spent so many minutes in a particular link. Alternatively, in a new experiment, imagine that a large number of people enter the maze (again the maze has no escape route). After sufficient wandering (sufficient for an equilibrium to obtain), the probability of a person's being in a particular link may be given as the ratio of the number of people in that link to the total number of people in the maze. Therefore, by taking an aerial picture of the maze, it is possible to say how much time a person would spend in each link had he or she wandered around the maze for a long time. The first case specifies the relative amount of time spent by one person in each link; the second case specifies the relative number of people in different parts of the maze in an instant of time after equilibrium prevails. The system is ergodic when these two probabilities are the same. In formal language, the statistical properties of a time series of a phenomenon (the individual maze-wanderer) are essentially the same as a set of observations made on a spatial ensemble (the spatial distribution of collective maze-wanderers) at a single time. In other words, ensemble averages can replace time averages in large-scale statistical statements. The individual maze-wanderer exemplifies the ergodic hypothesis, the collective maze-wanderers the ergodic theorem, which states that sampling across an ensemble (the people in the maze) is equivalent to sampling through time for a single system (the lone maze person).

How does this reasoning apply to geomorphology? One might discover that of all slopes in a region, 9 per cent stand at 6 degrees. If ergodic conditions apply, then the ergodic hypothesis would predict that the region would have 6-degree slopes for 9 per cent of its lifetime. In practice, few geomorphic applications of ergodicity make such quantitative statements about time using spatial data. This dearth of applications results largely from the strict conditions required for ergodic arguments to hold, including the difficulty of finding equilibrium landforms in an environment that is constantly changing. The few geomorphological applications that do meet the stringent statistical demands of the ergodic assumptions include studies of geomorphic magnitudes and frequencies, 'threshold' hillslopes, and the growth of drainage basins (see Paine 1985); river channel evolution (Zhang *et al.* 1999); and a general analogy between statistical thermodynamics and the transfer of mass within a landscape (Scheidegger 1991: 254).

A far commoner practice in geomorphology is to study change through time by identifying similar landforms of differing age at different locations, and then arranging them chronologically to create a time sequence or topographic chronosequence. Such space–time, or – more strictly – location–time, substitution has proved salutary in understanding landform development. Two broad types of location–time substitution are used. The first looks at equilibrium ('characteristic') landforms and the second looks at non-equilibrium ('relaxation') landforms.

In the first category of location–time substitution, the assumption is that the geomorphic processes and forms being considered are in equilibrium with landforms and environmental factors. For instance, modern rivers on the Great Plains display relationships between their width–depth ratio, sinuosity and suspended load, which aid the understanding of channel change through time (Schumm 1963). Allometric models are a special case of this kind of location–time substitution (see Church and Mark 1980).

Studies in the second category of location–time substitution, which look at developing or 'relaxation' landforms, bear little relationship to the ergodicity of physics. The argument runs that similar landforms of different ages occur in different places. A developmental sequence emerges by arranging the landforms in chronological order. The reliability of such location–time substitution depends upon the accuracy of the landform chronology. Least reliable are studies that simply assume a time sequence. Charles Darwin, investigating coral-reef formation, 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.

More useful are situations where, although an absolute chronology is unavailable, field observations enable geomorphologists to place the landforms in the correct order. This occasionally happens when, for instance, adjacent hillslopes become progressively removed from the action of fluvial or marine processes at their bases. This has happened along a segment of the South Wales coast, in the British Isles, where the Old Red Sandstone cliffs between Gilman Point and the Taf estuary have been affected by a sand spit growing from west to east (Savigear 1952). In consequence, the westernmost cliffs have been subject to subaerial denudation without waves cutting their bases the longest, while the cliffs to the east are progressively younger.

Relative-age chronosequences depend upon some temporal index that, though not fixing an absolute age of landforms, enables the establishment of an interval scale. For example, the basin hypsometric integral and stream order both measure the degree of fluvial landscape development and are surrogates of time (e.g. Schumm 1956).

The most informative examples of location–time substitution arise where absolute landform chronologies exist. Historical evidence of slope profiles along Port Hudson bluff, on the Mississippi River in Louisiana, southern USA, revealed a dated 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 stopping of undercutting. The changing conditions at the slope bases have reduced the mean slope angle from 40° to 22°.

Location–time substitution does have pitfalls. First, not all spatial differences are temporal differences because factors other than time exert an influence on landforms. Second, landforms of the same age might differ through historical accidents. Third, equifinality, the idea that different sets of processes may produce the same landform, may cloud interpretation. Fourth, process rates and their controls may have changed in the past, with human impacts presenting particular problems. Fifth, equilibrium conditions are unlikely to have endured for the timescales over which the locational data substitute for time, especially in areas subject to Pleistocene glaciations. Sixth, some ancient landforms are relicts of past environmental conditions and are in disequilibrium with present conditions.

Many geomorphologists substitute space for time to infer the nature of landform change. Only a handful of these adhere to the statistical assumptions of ergodicity. Nonetheless, the loose application of the ergodic reasoning, as seen in location–time substitution, is a productive line of geomorphological enquiry.

References

Brunsden, D. and Kesel, R.H. (1973) The evolution of the Mississippi River bluff in historic time, *Journal of Geology* 81, 576–597.

Church, M. and Mark, D.M. (1980) On size and scale in geomorphology, *Progress in Physical Geography* 4, 342–390.

Paine, A.D.M. (1985) 'Ergodic' reasoning in geomorphology: time for a review of the term? *Progress in Physical Geography* 9, 1–15.

Savigear, R.A.G. (1952) Some observations on slope development in South Wales, *Transactions of the Institute of British Geographers* 18, 31–51.

Scheidegger, A.E. (1991) *Theoretical Geomorphology*, 3rd completely revised edn, Berlin: Springer.

Schumm, S.A. (1956) Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey, *Bulletin of the Geological Society of America* 67, 597–646.

——(1963) Sinuosity of alluvial rivers on the Great Plains, *Geoglogical Society of America Bulletin* 74, 1,089–1,099.

Zhang, O., Jin, D. and Chen, H. (1999) An experimental study on temporal and spatial processes of wandering–braided river channel evolution, *International Journal of Sediment Research* 14, 31–38.

Further reading

Burt, T. and Goudie, A. (1994) Timing shape and shaping time, *Geography Review* 8, 25–29.

Thorn, C.E. (1988) *An Introduction to Theoretical Geomorphology*, 46–51, Boston: Unwin Hyman.

Thornes, J.B. and Brunsden, D. (1977) *Geomorphology and Time*, 19–27, London: Methuen.

RICHARD HUGGETT

ERODIBILITY

Erodibility is the resistance of surface material to erosion. It is usually restricted to soils or REGOLITH, and to water or WIND EROSION OF SOIL. Both water and wind erosion are complex processes, but when other factors are constant, erosion rates still vary due to differences in soil resistance. Erodibility is influenced by climate and is a complex, dynamic characteristic, which changes significantly over annual, seasonal or irregular time intervals, or even during a single storm. Nevertheless, the concept is useful for small-scale field or hillslope processes of rainsplash, sheetwash (see SHEET EROSION, SHEET FLOW, SHEET WASH) and rill erosion. It is more difficult to use with processes such as GULLY erosion, which involve very different spatial and temporal controls.

The erosion sub-processes affected by soil erodibility, are entrainment (by which soil particles are picked up), and transport. The relevant properties vary with the erosion process, and affect the erosive force available and resistance. In rainsplash erosion the entraining force is the kinetic energy of raindrop impact (see RAINDROP IMPACT, SPLASH AND WASH), converted to an upward force, while sheet wash and rill erosion involve runoff for both entrainment and transport. Raindrop impact can also disrupt particles and change their resistance to movement. The effectiveness of the upward force depends on soil particle size and mass. Poesen and Savat's (1981) experiments showed an entrainment:particle size relationship resembling the Hjulstrom curve for flowing water, particles between 0.1 and 0.25 mm diameter requiring least energy for movement. These results apply to non-cohesive particles with uniform density, such as quartz sands, where there is a direct relationship between particle size and mass. On such soils, erodibility can be assessed by standard particle size analysis techniques.

The relationship between particle size and erodibility is more complex when the surface is largely composed of aggregates. Aggregates are mixtures of mineral and organic matter, joined by electrostatic charges, microbial muscilages, hydrous oxides and carbonates. These materials and the volume of pore space are quite variable, so aggregate density is also variable. Microaggregate (<0.25 mm diameter) density is usually much higher than for macroaggregates, which may exceed 10 mm diameter (Oades 1993). Some large aggregates have densities below 1 g.cm^{-1}, can float on water, and are more easily eroded than small aggregates or mineral particles. The relationship between aggregate size and mass is not linear or direct, and size:entrainment relationships can be very complex.

These relationships are further complicated as aggregates can disintegrate during rainfall, if subjected to stresses exceeding the strength of 'stabilizing' agents. Under rainsplash, three dominant stresses occur. Raindrop energy, which can disrupt aggregates, is most significant, but SLAKING and differential hydration swelling may also cause breakdown. Aggregate strength is derived mainly from *cohesion*, due to electrostatic forces that bond clay and humus particles. Bond strength depends on total clay and humus content, on cation adsorption capacity, and on the cations adsorbed. Bivalent cations, such as calcium, cause flocculation and strong bonding, yielding small, highly resistant aggregates. Monovalent cations, such as sodium, cause particle repulsion, yielding weak, easily dispersed aggregates.

These interactions ensure that aggregates vary greatly in size, shape, stability and density, even in a single soil. As many soils consist largely of aggregates, it is their properties, rather than those of mineral particles, which most strongly influence erodibility. Aggregation is also affected by extrinsic factors such as wetting–drying and freeze–thaw, and by soil organic matter, changing with inputs of plant litter and decomposition by microbial activity. It is normal for regular or irregular seasonal changes in aggregation to be superimposed on short-term aggregate dynamics during or between storm events (Bryan 2000).

This discussion has emphasized the role of particles or aggregates as individual units. In fact, they only behave this way in recently disturbed or dispersive soils. The most common cause of soil disturbance is tillage. After tillage, disturbed soils gradually regain *coherence* due to weathering, compaction and crusting (see CRUSTING OF SOIL) by raindrop impact. This may occur in one rainstorm, or may take many months, but strongly affects erodibility. Erodibility usually declines as soils regain coherence, as sufficient force is required to overcome soil coherence and to entrain loosened particles.

Erodibility of coherent soils is determined by *soil shear strength*, the resistance to *interparticle failure*, defined by the Coulomb equation as:

$$r = c + zy \tan o$$

The active components are internal friction (o) which integrates surface and interlocking friction of particles, and (c) cohesion. Cohesion is explained above, while internal friction depends on the strength, heterogeneity of mineral particles and aggregates, and on overburden pressure (zy) at the point of potential failure. In surface soils acted on by rainsplash, sheetwash and rill erosion, overburden pressure is negligible and shear strength is dominated by cohesion. Both cohesion and internal friction are strongly influenced by soil water content. Cohesion is modified, either positively or negatively, by water molecules between particles. In completely dry soils, these are absent and soils are usually non-coherent and highly erodible. As soil water content increases, thin water films form (often < 1 micron in thickness) which are viscous and hold mineral particles together. As these become thicker, viscosity declines and cohesion is reduced. Internal friction also declines in wet soils, as positive pressure in water-filled pores counteracts overburden pressure. The strength of both coherent soils and soil aggregates is thus greatly reduced when soil water content approaches saturation, and erodibility increases.

The role of soil water means that erodibility also depends on soil hydrological properties. The overall control is climatic, but soil hydrological properties determine the proportion of water reaching the surface that infiltrates the soil, and how long the soil remains wet after a storm. Soil *infiltration capacity* depends on surface porosity, but on bare soils under intense rainfall, soil crusting or sealing often produces a thin almost impermeable surface layer. Once water infiltrates, its distribution and ultimate drainage depend on soil permeability, which is controlled by the *soil structure*, the arrangement of soil material and pore space, which determines such features as pore space continuity.

The properties described affect erodibility for all the processes discussed, but the relationship is somewhat different in each case, because of the role of surface water layers. Rainsplash often occurs without any surface water layer, and a significant water layer can reduce or eliminate erosive energy. In sheet wash and rill erosion, it is the *shear stress* exerted by runoff which causes entrainment and transport. As the existence and depth of the surface water layer is determined by the ratio of rainfall:infiltration capacity, this means that for sheetwash and rill erosion, shear stress is partially controlled by soil properties. The magnitude of shear stress exerted by runoff is also affected by water distribution, increasing significantly when runoff is concentrated on a small surface area. Flow concentration depends on surface roughness, which is controlled by vegetation and on bare surfaces by soil particle and aggregate heterogeneity. These properties determine whether, under raindrop impact, the surface becomes progressively rougher or smoother, and will therefore affect the distribution and magnitude of runoff shear stress.

The complexity of the interacting processes and properties involved means that erodibility is not a single, simple, measurable soil property (Lal 1990), but reflects the collective interaction of many soil properties. Nevertheless, it is often necessary to attempt to assess erodibility by one or several simple measures. Many attempts have been made to identify or combine soil properties as *indices of soil erodibility* (Bryan 1968). No single measure is successful in all cases, but several measures can be effective if the precise nature of the dominant erosion process is clearly identified. Measures of aggregate water stability, such as wet-sieving, are useful for rainsplash erosion, particularly on disturbed soils, while soil shear strength measured with a *vane shear* apparatus can be useful on coherent soils. *Soil consistency*, assessed by Atterberg limits, can indicate crusting potential. As sheetwash and rill erosion are more complex processes, it is more difficult to isolate a reliable index, but measures based on range of aggregate or particle size may be promising. In all cases, however, the high temporal and spatial variability of erodibility must be recognized.

References

Bryan, R.B. (1968) The development, use and efficiency of indices of soil erodibility, *Geoderma* 2, 5–26.

——(2000) Soil erodibility and processes of water erosion on hillslopes, *Geomorphology* 32, 385–415.

Lal, R. (1990) *Soil Erosion in the Tropics: Principles and Management*, New York: McGraw-Hill.

Morgan, R.P.C. (1995) *Soil Erosion and Conservation*, London: Longman.

Oades, J.M. (1993) The role of biology in the formation, stabilization and degradation of soil structure, *Geoderma* 56, 377–400.

Poesen, J. and Savat, J. (1981) Detachment and transportation of loose sand by raindrop splash, Part II: Detachability and transportability measurements, *Catena* 8, 19–41.

RORKE BRYAN

EROSION

Commonly speaking the term erosion (from Latin *erodere* = to gnaw away) is often used to indicate the overall exogenic process or group of processes that are directed at levelling off Earth relief, in contrast with the antagonist endogenic processes (crustal movements and volcanism) that build it up. In this very wide meaning erosion includes: acquiring materials from the higher elevations, moving them from one place to another (*transport*) and leaving them in lowlands (*deposition*).

Actually, it is the opinion of all scientists that erosion cannot include deposition. In fact, in a more technical language, the term erosion – although variously defined – usually excludes the processes whereby transported materials are set down. In the most broad and common of the technical meanings, erosion includes all exogenic processes, excluding WEATHERING and MASS MOVEMENTS, that involve the entrainment of loose weathered materials by a mobile agent, the removal of bedrock particles by the impact of transported materials, the mutual wear of rock fragments in transit and the transportation of acquired materials (Thornbury 1954).

Sometimes the term is restricted by excluding transportation; in this case DENUDATION is the more general term. More rarely, erosion indicates exclusively the entrainment of loose materials by a mobile agent.

Erosional agents and processes

Erosional processes are performed by mobile agents that draw their energy from solar radiation and act in one or more ways, constantly driven by the force of gravity. The principal erosional agents are running water, glaciers, wind and sea waves (Figure 57). In some cases they complete the same process, in some others a given process is accomplished by a distinctive agency that operates according to its physical peculiarities. Beside the cited 'natural' agents, humans can be rightly considered an important erosional agent too. Nowadays anthropogenic activities are so widespread and marked that they deeply modify the Earth's surface, often in an irreversible manner (see ANTHROPOGEOMORPHOLOGY).

The most common processes performed by natural erosional agents are shown in Table 17. The entrainment of rocky particles by erosional agents can be both chemical and mechanical. The first action (CORROSION) implies the work of a solvent and therefore it is typically accomplished by running water or waves; it is less important than mechanical action.

Mechanical removal takes place with different modalities, depending on the erosional agent. *Hydraulic action* comes from the pressure and hydraulic force of flowing water or sea waves that allow the acquisition of rocky particles. CAVITATION is a particular process operated by running water; it is still poorly documented and would represent a mechanism through which hydraulic action has a direct role in bedrock breakage. This process occurs when an increase in flow velocity and the following decrease in pressure cause the formation of bubbles that implode emitting jets of water capable of fracturing solid rocks. Moving ice acquires materials by *plucking* (or *quarrying*); through this process glaciers that move forward can remove large rocky fragments already detached from the bedrock by the freezing of water circulating inside cracks. *Overdeepening* is the process whereby glaciers erode to levels below regional BASE LEVEL related to fluvial systems;

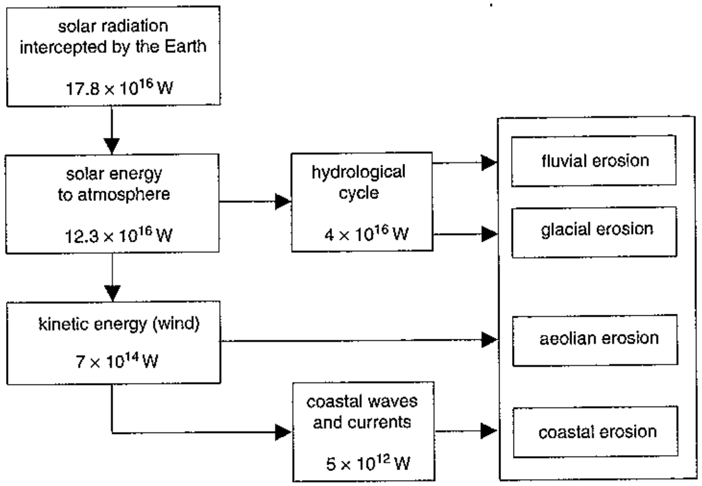


Figure 57 Source and flow of energy available for the different kinds of erosional processes. (From Summerfield 1991: 21, simplified and modified.)

Table 17 Erosional agents and their relevant erosional processes

Erosional Agent	Erosional processes			
	Entrainment of rocky materials	Erosion by transported materials	Wear of transported materials	Methods of transportation
Running water	Hydraulic action (Corrosion)	Abrasion	Attrition	Traction Suspension Solution
Glacier	Plucking or quarrying	Abrasion	Attrition	Traction
Wind	Deflation	Corrosion or abrasion	Attrition	Traction Suspension
Waves and currents	Hydraulic action (Corrosion)	Abrasion	Attrition	Traction Suspension (Solution)

Source: From Thornbury (1954: 47, modified)
Note: The less effective processes are indicated in brackets

however, it would be more correct to consider overdeepening as one of the effects of glacial erosion rather than an erosional process (Castiglioni 1979). The turbulent eddy action of wind is responsible for *deflation* and produces effects that are similar to those deriving from hydraulic action of waters.

Erosion accomplished by transported materials (named ABRASION) is due to the continuous collisions and friction by particles in transit on bedrock. All erosional mobile agents carry out this process. Abrasion due to running water is operated by solid materials of any size (up to large boulders, depending on the flow velocity) that can

be transported as BEDLOAD by the water flow. Breaking waves that throw solid particles against the shore perform the same abrasive effect. EVORSION is a particular kind of abrasion due to the action of running water. It is caused by the erosional action of vortices and eddies on stream rocky beds. When a stationary eddy rotates a pebble, a small hollow is produced; this process leads to the formation of POT-HOLES (evorsion hollows), which contribute to valley deepening. In glacial environments abrasion is the friction produced on the bedrock by the debris carried along in the basal parts of glaciers; in its broader meaning it can include STRIATION, i.e. the bedrock *scoring* and *polishing* that reduces the rock surface roughness (Benn and Evans 1998). Aeolian abrasion derives from the repeated impacts of sand grains, silt particles and dust on rock surfaces; it is more properly named *corrasion*.

The wear of transported solid particles (*attrition*) takes place through repeated reciprocal knocking and collisions among the materials in transit: the result is a progressive decrease of particle sizes. At the same time the rocky particles tend to assume particular shapes, depending on the different ways in which each agent accomplishes transportation.

Mobile agents carry out transportation in three different ways. *Traction* consists of the rolling, sliding, pushing or jumping (in which case SALTATION is the specific term) of transported particles that are swept along on or immediately above a bottom surface. *Suspension* is a mode of transportation by water and wind; it implies the holding up of transported particles by the upward currents that develop in turbulent flows like those of running water and moving air (see SUSPENDED LOAD). *Solution* is a kind of 'chemical' transportation that is restricted to the action of running water and waves.

Erosional processes produce distinctive erosional landforms; furthermore, each erosional agent develops its own typical assemblage of landforms, depending on its mode of shaping Earth's relief. Erosional landforms are particularly striking features of the landscape; for this reason, perhaps, they have been considered with more attention than depositional landforms that, although interesting from a morphogenetic point of view, are usually less attractive.

The close links between erosional agency, accomplished process and produced landforms were recognized also by the ancients. Leonardo da Vinci at the end of the fifteenth century wrote: 'Every valley is created by its river and the proportion between one valley and another is the same as that between one river and another.' Once the concept that distinguishing features of landforms depend on the geomorphic process responsible for their development was fully understood, the genetic classification of landforms became possible. This scientific advancement made it possible to study the Earth's physical landscape not only from the descriptive point of view, but also considering the possible interpretation as to its geomorphological history. The genetic interpretation of erosional landforms, however, to be satisfactory must take into account the possible homologous or converging landforms, i.e. those landforms that although generated by different processes show similar features. Moreover it must be kept in mind that the genesis of most erosional landforms cannot be attributed to a single process, although it is rather simple to identify the dominant one.

The work of erosional agents produces peculiar assemblages of landforms that take on distinctive aspects depending on the stage of their development. The recognition that landforms change in time sequentially is the basis of the concept of the *Geographical Cycle* by Davis (1899). Once this concept had imposed itself, geomorphological interpretations made a new step forward. In fact, if properly applied, the geographical cycle affords a useful reference scheme to predict the possible future evolution of Earth's physical landscape (see CYCLE OF EROSION).

Erosion factors

Erosional landform features strictly depend not only on the way the exogenic agents operate, but also on a series of factors that control both the nature and the rate of erosion. The most important factors of erosion are lithology, tectonics, climate, vegetation and humans.

LITHOLOGY

Lithology strongly controls erosional processes as rock ERODIBILITY relies on it; as a consequence it influences the speed of erosional processes. In this perspective rocks are often referred to as 'hard' or 'resistant' or 'weak' and 'non-resistant' to erosional processes. The same erosional process can operate in a differentiated way where resistant rocks crop out next to no resistant rocks: as the erosional process proceeds, an uneven surface originates where more resistant rocks, slowly and hardly eroded, stand higher above less resistant rocks, which are more quickly and easily eroded. To some extent differential erosion can produce INVERTED RELIEF. The effects of differential erosion are particularly evident on stratified and differently erodible rocks. In this case the result of erosion is the formation of steep, abrupt faces of rock that mark the outcrop of the more resistant layers; the steep faces of a CUESTA, the rock terraces of a step-like slope or the scarp of a MESA are typical products of differential erosion. The concept of more or less erodible rocks is a relative one; in fact a rock can be resistant to one process and non-resistant to another. Therefore lithology has an influence also on the typology of the erosional processes.

TECTONICS

Tectonics influences erosional processes in different ways. Faults (see FAULT AND FAULT SCARP) and FOLDs can bring into contact rocks with different erodibility and then favour differential erosion. Furthermore they can directly influence the response of rocks to erosion, thus conditioning erosion rate. In fact rock erodibility depends not only on lithological characteristics but also on rock attitude (dip-slopes are less resistant than scarp-slopes) and on the degree of tectonic deformation (the higher the deformation of rocks, the higher their erodibility). Tectonic joints and faults can influence both the intensity of erosion and the location of the resulting landforms (see JOINTING). For example, fluvial erosion acts more powerfully where joints and faults create zones of weakness in the rocks than in other directions. As a consequence the orientation of stream valleys often follows closely the directions of these discontinuities. The sensitivity of tectonic discontinuities towards erosional processes may be so great that the morphological effects of differential erosion can help in the identification of discontinuities of small entity or affecting plastic lithologies (Belisario *et al.* 1999).

Tectonic uplift has also an important role in controlling the effectiveness of erosional processes. Uplift and erosion, together with relief, are the fundamental components of geomorphodynamic systems (see SYSTEMS IN GEOMORPHOLOGY) and are functionally related to one another in a negative feedback process (Ahnert 1998). When uplift prevails relief increases, and as a consequence erosion rate is faster. Increased erosion rate can eventually balance the building processes; in this case mountains do not change in elevation. When erosion overcomes the effects of uplift, elevations begin to lower and consequently the erosion rate slows down until the whole process comes to an end (Figure 58).

CLIMATE

Climate controls erosional processes both directly and indirectly. The direct control is exerted by the climate elements – temperature, rainfall and wind – that show a wide variability not only from one part to another of our planet, but also within very restricted areas, as for example from one slope to another on the same mountain. This wide variability of climatic conditions affects WEATHERING processes that weaken the rocks, predisposing them to subsequent erosional processes. Furthermore it favours the action of one erosional agent with respect to others: fluvial erosional processes become dominant in shaping the Earth's surface where rainfall amount is enough to guarantee the perennial channelled flow of waters, wind erosion is particularly effective where humidity is low, and glaciers can operate only where temperatures are such as to allow the fall and accumulation of snow. Besides this quite obvious influence, climatic conditions also control the way the different erosional processes operate with each other; these considerations are the basis of CLIMATIC GEOMORPHOLOGY which examines the systems of morphodynamic processes and their reciprocal interactions, in relation to the different climatic conditions.

Climate not only affects the typology of erosional processes but also the different behaviour of rocks. Under different climatic conditions the same rock may exhibit a different degree of resistance towards erosional processes and therefore it can be shaped into a variety of landforms. Granitic rocks are a good example. Depending on the climatic conditions and therefore on the dominant erosional processes they can be eroded into: the sharp peaks of the Monte Bianco massif, the low relief of INSELBERGs that dominate the savannah and steppe regions of South Africa, the ellipsoidal hollows (TAFONI) which originate from chemical corrosion and from the sweeping action of wind, the large round-topped mountains (piton) of tropical regions, etc. The close relationships among climatic conditions, erosional processes and landforms help to reconstruct the climatic variations that occurred in the past by

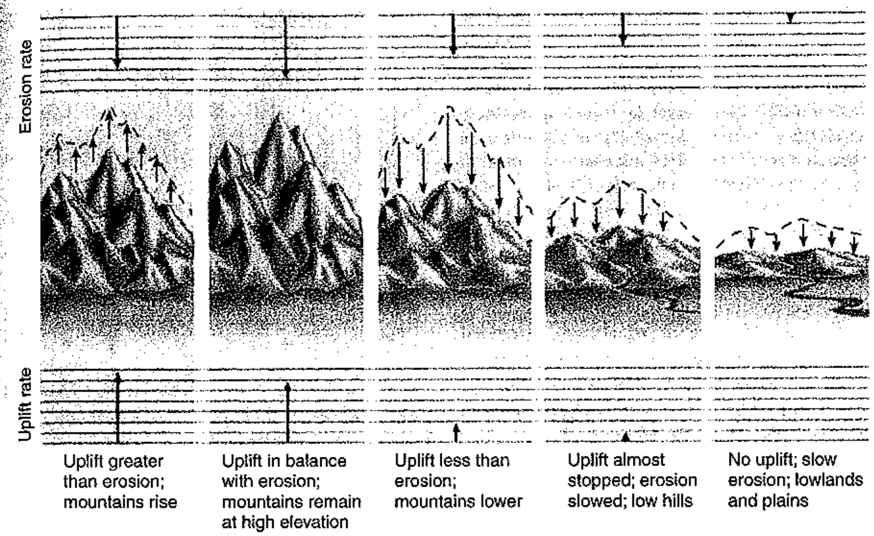


Figure 58 Negative feedback process relating uplift, erosion and mountain elevation. (From Press and Siever 1994: 364, modified)

examining the imprints left on the Earth's landscape by the prevailing erosional processes.

VEGETATION

The indirect influence of climate on erosion is largely related to the way it affects the amount and type of vegetation that, in its turn, has an important control on the EROSIVITY of some erosional agents. A thick vegetation cover inhibits surface runoff, thus restraining the action of running water; furthermore it obstructs the free flow of winds and therefore reduces the effectiveness of aeolian erosional processes. Root structures have a double influence: they can enhance the resistance of loose materials towards erosion or they can cause the breakage of solid bedrock, thus making erosion easier. As a whole, vegetation constrains erosional processes more frequently than it favours them. The limiting effect of vegetation on erosional processes varies as a function of the kind and density of the vegetation cover, and reveals itself because it enhances the stability of surface materials. Vegetation is also an important factor of pedogenesis, as it supplies the organic matter necessary for humus formation; therefore it has a role both in forming soil and in protecting it from erosion. This protective action is of primary importance to inhibit SOIL EROSION. When climatic conditions are such as to assure a dense and persisting vegetation cover erosional processes are slowed down: in these conditions (referred to as biorhexistasy) soils can develop and stay in place; on the contrary when climate is unfavourable to the development of vegetation, erosional processes are largely widespread and soils are easily removed (rhexistasy).

HUMAN IMPACT

If it is true that *humans* are nowadays powerful erosional agents, they are also an important factor of erosion. Anthropogenic activities are sometimes directed to undo or reduce erosional processes accomplished by 'natural' agents, as, for example, in the case of coastal defences built to inhibit sea erosion. More frequently, however, they produce the opposite effect and make erosion rate faster: in very densely inhabited areas, for instance, the extensive use of asphalt and concrete favour surface runoff and then erosion due to running water.

All the erosion-controlling factors play their role together; therefore their effects can interfere with each other in many possible ways so that the overall control on erosional processes is widely differentiated both in space and time. An example that clarifies the response of erosional processes to the complex constraints imposed by erosional factors is afforded by some more careful considerations about soil erosion. Once pedogenetic processes have led to formation of soils, they are exposed to the action of exogenetic erosional agents that start consuming them. When soil erosion proceeds normally, equilibrium conditions are attained: the rate at which soil is eroded equals the rate of soil formation. If this equilibrium is broken, erosional processes can become faster than pedogenetic processes; as a result, soil erosion is accelerated. The conditions more favourable to start *accelerated erosion* occur where weak rocks (such as clay or marls) crop out on areas affected by abundant and irregular precipitation that favours erosion by running water; under these conditions, for instance, BADLANDS originate. Wherever these natural predisposing conditions are added to deforestation and faulty land use connected to anthropogenic activities, accelerated erosion attains its maximum intensity. Under these conditions the soil erosion rate exceeds the soil formation rate. As a result soils get thinner and can completely disappear. In some cases erosion becomes so severe that it can be compared to the process of DESERTIFICATION: the irreversible process whereby soils lose their fertility because of the destructive effects of some anthropogenic activities.

References

Ahnert, F. (1998) *Introduction to Geomorphology*, London: Arnold.
Belisario, F., Del Monte, M., Fredi, P., Funiciello, R., Lupia-Palmieri, E. and Salvini, F. (1999) Azimuthal analysis of stream orientations to define regional tectonic lines, *Zeitschrift für Geomorphologie, Supplementband, NF* 118, 41–63.
Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.
Castiglioni, G.B. (1979) *Geomorfologia*, Torino: UTET.
Davis, W.M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.
Press, F. and Siever, R. (1994) *Understanding Earth*, New York: Freeman and Company.
Summerfield, M.A. (1991) *Global Geomorphology*, Harlow: Longman.
Thornbury, W.D. (1954) *Principles of Geomorphology*, New York: Wiley.

Further reading

Fournier, F. (1960) *Climat et érosion la relation entre l'érosion du sol par l'eau et les précipitations atmosphériques*, Paris: Presses Univ. de France.
Haggett, P. (1983) *Geography. A modern Synthesis* (Chapter 10), New York: Harper & Row.
Howard, J.A. and Mitchell, C.W. (1985) *Phytogeomorphology*, New York: Wiley.
Keller, E.A. and Pinter, N. (1996) *Active Tectonics, Earthquakes, Uplift and Landscape*, New Jersey: Prentice Hall.
Scheidegger, A.E. (1979) The principle of antagonism in the Earth's evolution, *Tectonophysics* 55, T7–T10.

SEE ALSO: fluvial geomorphology; glacial erosion; granite geomorphology; tectonic geomorphology; wind erosion of soil

ELVIDIO LUPIA-PALMIERI

EROSIVITY

A measure of the potential ability of a soil to be eroded by particular geomorphological processes. Erosion is a function of erosivity on the one hand and of erodibiltiy (the vulnerability of a soil to erosion) on the other.

Water erosion susceptibility is related to various rainfall erosivity indices. Rainfall intensity, rainfall amount and antecedent conditions are all important controls of erosivity, but as Morgan (1995: 27) has remarked: 'the most suitable expression of the erosivity of rainfall is an index based on the kinetic energy of the rain. Thus the erosivity of a rain storm is a function of its intensity and duration, and of the mass, diameter and velocity of the raindrops'. In recent years RAINFALL SIMULATION has been used to assess the response of soils to storms with different characteristics.

Wind erosion susceptibility has also often been determined using indices based on wind velocities and durations above certain threshold velocities (e.g. Skidmore and Woodruff 1968) and portable wind tunnels have been employed to assess the response of different ground surfaces to different wind velocities.

References

Morgan, R.P.C. (1995) *Soil Erosion and Conservation*, 2nd edition, Harlow: Longman.
Skidmore, E.L. and Woodruff, N.P. (1968) Wind erosion forces in the United States and their use in predicting soil loss, *USDA Agricultural Research Service Handbook* 346.

A.S. GOUDIE

ERRATIC

Erratics are rock fragments carried by a glacier, or in some cases by floating ice, and subsequently deposited at some distance from the outcrop from which they were derived. For this reason their lithology differs from the surrounding rocks and sediments – hence the term 'erratic'. Some erratics are large blocks, that lie free on the surface and form interesting landscape features. Glaciologists, however, mainly use the term for the exotic components embedded in tills (see GLACIAL DEPOSITION), encompassing both large clasts and fine-grained rock fragments.

Scientific investigation of erratics started during the first half of the nineteenth century, when most geologists thought that they were swept into the northlands by the universal flood. At the same time their transportation and distribution by former widespread glaciers was first suggested, and later in the nineteenth century it was universally agreed.

Some erratics form landmarks because of their spectacular dimensions. One of the largest erratics measures 45 m by 20 m by 10 m and is estimated to weigh 16,500 tons. It is part of the Foothills erratic train, a series of boulders stretching over 400 km along the eastern foothills of the Rocky Mountains.

Erratics not only give evidence of the existence of former glaciers; especially erratics in tills provide a powerful tool for many glacial investigations. Such studies are based on the identification of 'indicator erratics'. Indicator erratics are those for which a definite source area is known. They form 'indicator trains' or, in cases of shifting ice divides and ice flow directions within an ice sheet, 'indicator fans' trailing downglacier from the source rock. Indicator trains and fans are enriched in the distinctive component relative to the till underlying or enclosing it. The concentrations of indicator erratics vary systematically along former ice flowlines. Within indicator outcrops, concentrations increase rapidly downglacier, reflecting the addition of new material from the glacier bed, but concentrations drop off rapidly down-ice of the outcrop margin. The up-ice and down-ice limits of an indicator plume are known as the 'head' and the 'tail', respectively.

Erratics can be used to reconstruct the pattern and history of ice flow in studies of ice sheet dynamics as long as erratic transport histories are not blurred by repeated glaciations involving total redistribution of previously deposited materials (Benn and Evans 1998). The study of erratic dispersal patterns furthermore can provide important clues to the location of mineral outcrops or ore bodies, because the erratic plumes are much larger than their bedrock sources, making them easier targets to find (Kujansuu and Saarnisto 1990). In Denmark, Germany and the Netherlands till units of different age show differently composed indicator assemblages and counts of the erratics derived from various western, central or eastern Scandinavian source areas are here successfully employed in stratigraphical investigations (Ehlers 1996).

References

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.
Ehlers, J. (1996) *Quaternary and Glacial Geology*, Chichester: Wiley.
Kujansuu, R. and Saarnisto, M. (eds) (1990) *Handbook of Glacial Indicator Tracing*, Rotterdam: Balkema.

CHRISTINE EMBLETON-HAMANN

ESCARPMENT

The term escarpment, or scarp, has been applied traditionally to a steep, often single slope, of considerable length, that dominates a section of landscape. An escarpment thus can be distinguished from the two flanking walls of canyons. For example, south of Sydney, Australia, the coastal escarpment forms a long, virtually continuous wall, but it is outflanked by canyons which extend more than 100 km further inland. Another notable instance of a valley cut well back from an escarpment is the Sognefjord, which extends about 200 km inland from the coastal edge of the Norwegian highlands. The lengths of escarpments vary from a few kilometres to the subcontinental scale of mega-escarpments, or Great Escarpments, such as the Drakensberg of South Africa, while their heights vary from a few tens of metres to several thousand metres. A distinction is generally drawn between denudational escarpments and fault scarps (see FAULT AND FAULT SCARP), although this may be no simple exercise in areas of essentially homogeneous crystalline rocks.

The majority of denudational escarpments have formed as a result of differential rock resistance to erosion. Probably the most outstanding examples form the sequence of the Vermillion, White, Grey and Pink Cliffs, known as Great Staircase, which rises from the rim of the Grand Canyon of the Colorado River to the high plateaux of southern Utah. The treads of this staircase are cut mainly in softer strata between the sandstones which form the cliffs. Major sandstone escarpments also occur in the Adrar and Borkou areas of the Sahara. But such features are not limited to arid lands, for the great ramparts of the Roraima massif have developed in the humid tropics of Venezuela. Neither are they limited to sequences of sedimentary rocks. Major escarpments and benches also have developed in response to the differential resistance of volcanic strata in the Bushveld of Transvaal, and of sheeted granites intruded through metamorphic rocks in Madagascar. Some escarpments in granitic rocks were initiated at the boundary with less resistant rocks, and have retreated to their present positions. Others, such as the multiple steps in the Sierra Nevada of California, were initiated by differential weathering within a granitic mass.

While many escarpments can be attributed to the great resistance to erosion of particular types of rocks, others show no systematic relationship to lithology. For example, the coastal escarpment south of Sydney extends from sedimentary to metamorphic rocks, and thence to crystalline rocks. Escarpments of this type have developed in response to regional uplift. The most extensive of them, which are sometimes called Great Escarpments, occur on many tectonically PASSIVE MARGINs of continents. Notable examples are the Drakensberg, the Western Ghats of India, the escarpments of east Australia, the Serra da Mantiqueira of Brazil, the coastal escarpment of Norway, the escarpment of east Greenland and the Torngat Mountains of the uplifted margin of Labrador.

Great Escarpments are not limited to passive margins, for they also occur on tectonically active continental margins. Collision of the Australian and Pacific Plates has resulted in some 20 km of uplift in about the last 10 million years along the west coast of the South Island of New Zealand. And, although uplift has been virtually matched by erosion, the flank of the Southern Alps rises in a steep, heavily dissected wall from the narrow coastal plain. Likewise, collision of the Pacific, North American and Cocos Plates has resulted in major uplift that, together with rapid erosion, has produced the great escarpments of the Sierra Madre Occidental and Sierra Madre Del Sur on the western flank of the Mexican highlands.

Less extensive, though nonetheless impressive, escarpments have developed as a result of regional uplift in continental interiors. The classic examples are along the margins of the *Mittelgebirge* of central Europe, such as the Massif Centrale of France and the Erzgebirge of Germany. However, even more impressive escarpments occur along the margins of uplands in central Asia, as for example on the northern side of the Bogda Shan in north-west China. Erosion in response to the regional uplift in north-east Africa has resulted in high escarpments cut largely in volcanic rocks along the west flank of the Ethiopian plateau. The Kaibab Limestone escarpment of the Mogollon Rim on the southern flank of the Colorado Plateau is a notable North American example.

Much of the initial research on escarpments was carried out in the folded terrain of the Appalachians and north-west Europe, where they occur in association with homoclinal CUESTAs and HOGBACKs. As early as 1895 W.M. Davis pointed out escarpments in this type of terrain were second-stage features that did not develop until streams began to extend headwards along the strike of the folds. He attributed the retreat of escarpments not only to erosion on their steep faces, but also to lateral erosion of streams at the foot of escarpments. Although Davis's ideas, and especially his terminology, have been subject to much criticism (see SLOPE, EVOLUTION), the scheme that he proposed a century ago still provides the basis for the interpretation of many scarp and cuesta landscapes. However, major challenges to it have come from German geomorphologists.

Schmitthenner (1920) argued that the most important form of denudation in scarplands (*Schichtstufenlandschaft*) is the lowering of the backslope surface and the breaching of escarpments from the rear. He attributed this primarily to seepage down the dip, and the consequent development of swampy hollows, or *dellen*, by solutional processes. Strong support for the role of seepage leading to the lowering of escarpment crests and to the breaching of them from the rear has come from recent research on the Colorado Plateau and Australia.

Many stairways of multiple escarpments have been attributed to repeated uplift and erosion, but, as independent evidence of repeated uplift is often lacking, alternate hypotheses need consideration. W. Penck (1924) claimed that multiple scarps and benches (*Piedmonttreppen*) could form on a continuously rising and expanding dome. Penck's hypothesis provides a valuable warning that the relationship between tectonics and slope form may be complex (see SLOPE, EVOLUTION).

In recent decades prominent German researchers have expounded climatic explanations of escarpments (see MORPHOGENETIC REGION). Büdel (1982) argued that escarpments in the tropics are essentially the result of deep weathering and the subsequent stripping of regolith, and referred to them as 'etchplain stairways'. He also extended this climatic interpretation to temperate lands by claiming that most of the escarpments of south-west Germany were formed in a similar fashion to etchplain stairways under a past 'tropicoid climate'. The structural influences so clearly expressed in many of these escarpments were dismissed as only an 'arabesque' in the general two-stage development by deep weathering and subsequent stripping. According to this theory, scarplands are sculpted predominantly by areal downwearing, and scarp retreat is minimal.

In striking contrast, conclusions drawn from studies in southern Africa, especially those of L.C. King, emphasize the dominant role of the retreat of escarpments over long distances. According to King (1953), the most important processes are sheet wash on pediments at the foot of scarps, and mass failure and gully erosion on the steep slopes. He argued that 'scarps retreat virtually as fast as nick-point advance up rivers, so that the distribution of successive erosion cycles bears no relationship to the drainage pattern whatsoever'. He argued also that retreating scarps resulted in isostatic (see ISOSTASY) uplift, generally in the form of large-scale warping, initiating a new cycle of scarp retreat.

Although many escarpments may have retreated over considerable distances, some have apparently not done so. The Blue Ridge Escarpment on the east flank of the Appalachian Mountains is a major feature, which lies about 350 km inland from the coast. Rather than having retreated from the coast, however, this escarpment may have been maintained in approximately its present position over many millions of years by uplift along an ancient continental margin preserved deep in the crust. It thus seems to be in a long-term state of dynamic equilibrium in which denudation has been largely balanced by the slow rise of the underlying rocks. On the other hand, some escarpments are essentially fossil features that have been exhumed by erosion (Plate 41). Sedimentary evidence indicates that the Arnhemland Escarpment, cut in Proterozoic rocks in the Northern Territory of Australia, is a coastal cliff-line that was buried during Cretaceous times and subsequently exhumed. Since being exhumed, parts of this escarpment have been almost stationary, and the most active parts have retreated only a few kilometres.

Assessing which explanation is best suited to a particular escarpment may be no simple task, and indeed none of them may be entirely satisfactory. The final appeal must be to field evidence rather than to conceptual constructs.

The rates at which escarpments retreat vary greatly, and seem to depend on lithological resistance, tectonic activity and the intensity of erosion which itself is largely controlled by climate. Many major escarpments are thought to retreat about 1 km per million years. However, the distance of retreat from dated basalts show that escarpments in temperate east Australia have retreated at rates of only about 25 to 170 m per million years, and the rates for some of them have been even slower. Low rates of only about 70 m to 150 m per million years have also been recorded in subarctic Spitzbergen. Average rates of scarp retreat on the Colorado Plateau since Miocene times have been only about 160 m per million years, but earlier than that the rates were about 1.5 km to 4 km per million years. The great reduction has been attributed to the incision of streams and thus the cessation of very active lateral planation at the foot of scarps. Denudation

Plate 41 Proterozoic sandstone capping schists on the Arnhemland Escarpment, Northern Territory, Australia

on the Arnhemland Escarpment increased by an order of magnitude during the late Quaternary, but in this case the change seems to have been climatically controlled. Moreover, the evidence from Arnhemland prompts caution in extrapolating long-term rates of retreat from relatively short-term records.

Although most research has been carried out on escarpments above sea level, submarine escarpments are of far greater magnitude. The fall below the continental shelf off east Australia is more than five times greater than the height of the escarpment onshore, and contrasts of similar, or greater, magnitude occur along most continental margins.

References

Büdel, J. (1982) *Climatic Geomorphology*, Princeton: Princeton University Press.

Davis, W.M. (1895) The development of certain English rivers, *Geographical Journal* 5, 127–146.

King, L.C. (1953) Canons of landscape evolution, *Geological Society of America Bulletin* 64, 721–752.

Penck, W. (1924) *Die Morphologiche Analyse; ein Kapital der Physikalischen Geologie*, Stuttgart: Engelhorn. Trans H. Czech and K.C. Boswell (1954) *Morphological Analysis of Landforms*, London: Macmillan.

Schmitthenner, H. (1920) Die Enstehung der Stufenlandschaft, *Geographische Zeitschrift* 26, 207–229.

Further reading

Battiau-Queney, Y. (1988) Long term landform development of the Appalachian piedmont (USA), *Geografiska Annaler* 70A, 369–374.

Blume, H. (1958) Das morphologische Werk Heinrich Schmitthenners, *Zeitschrift für Geomorphologie* 2, 149–164.

Nott, J.F. (1995) The antiquity of landscapes on the Northern Australian craton; implications for models of longterm landscape evolution, *Journal of Geology* 102, 19–32.

Ollier, C.D. (1985) Morphotectonics of passive continental margins: introduction, *Zeitschrift für Geomorphologie Supplementband* 54, 1–9.

Young, R.A. (1985) Geomorphic evolution of the Colorado Plateau margin in west-central Arizona: a tectonic model to distinguish between the causes of rapid symmetrical scarp retreat and scarp dissection, in J.T. Hack and M. Morisawa (eds) *Tectonic Geomorphology*, 261–278, London: Allen and Unwin.

Young, R.W. and Wray, R.A.L. (2000) Contribution to the theory of scarpland development from observations in central Queensland, Australia, *Journal of Geology* 108, 705–719.

R.W. YOUNG

ESKER

Derived from the Irish word *eiscir* (ridge), an esker is an elongate sinuous ridge composed of glacifluvial sediments and marking the former position of a subglacial, englacial or supraglacial stream. The routing of former meltwater channels in glaciers, and their association with ice-marginal landforms and sediments is indicated by the overall form of eskers. There are four major types of esker (Warren and Ashley 1994: (1) continuous ridges (single or multiple) that document tunnel fills; (2) ice-channel fills produced by the infilling of supraglacial channels; (3) segmented ridges deposited in tunnels during pulsed glacier recession; and (4) beaded eskers consisting of successive subaqueous fans deposited in ice-contact lakes during pulsed glacier recession. In planform, eskers also come in a wide variety of types. These include, single ridges of uniform cross section or of variable volume, single ridges linking numerous mounds (beads) and complex braided or anabranched systems. Although individual eskers or esker networks may stretch over hundreds of kilometres of former ice sheet beds it is unlikely that they formed in tunnel systems of that length. Rather, they were probably deposited in segments in the ice sheet marginal zone of ablation and each segment was added as the ablation zone migrated towards the ice sheet centre. In some locations eskers lie on the bottoms of bedrock meltwater channels (TUNNEL VALLEYS or Nye channels) indicating that erosion by subglacial meltwater was followed by deposition, possibly due to waning discharges, and that subglacial conduits were remarkably stable features.

Most eskers are aligned sub-parallel to glacier flow, reflecting the flow of meltwater towards the glacier margin. Eskers that were deposited as subglacial tunnel fills may be the result of flow in pressurized conduits and therefore may possess up-and-down long profiles where they climb over topographic obstacles. This is due to the fact that flow in the conduits was driven by the ice surface slope rather than by the glacier bed topography (Shreve 1972). If conduits or tunnels switch to atmospheric pressure, as occurs beneath the thinner ice nearer to the glacier snout, then the water will follow the local bed slope and so any resulting eskers will be deposited transverse to glacier flow (*valley eskers*).

The former englacial position of some eskers is indicated by the occurrence of buried glacier ice or almost complete disturbance of the stratified core. Englacial or supraglacial construction of eskers will result in the draping of the features over former subglacial topography after glacier melting. These apparent up-and-down long profiles must not be confused with true subglacial eskers deposited under pressure and therefore also crossing topographic high points. Largely intact internal stratigraphies are typical of subglacial eskers.

Eskers are often well stratified but contain a variety of sediment facies. Particles are usually not far-travelled, most being no further than 15 km from their source outcrop. The wide range of BEDFORMS observed in esker sediments reflect and document the large variations in meltwater discharge on both diurnal and seasonal timescales. Rhythmicity or cyclicity in the sediments is often represented by fining-upwards sequences separated by erosional contacts. Each fining-upward unit records an individual discharge event of maybe only hours in duration. The occurrence of large cross-bedded sequences may document deposition in deltas in subglacial ponds. Where tunnels collapse and/or change shape or streams change size or position, one depositional sequence may be truncated or partially infilled by another. Apparent anticlinal bedding in some eskers is thought to be the result of sediment slumping down the esker flanks as the supporting ice walls melt back.

The segments or beads on eskers are interpreted as the products of ice marginal deposition at the mouths of subglacial tunnels. At each tunnel mouth the glacifluvial sediment being carried through the tunnel or conduit is deposited in SUBAERIAL or subaqueous fans/deltas due to the sudden drop in meltwater velocity as it leaves the confines of its ice walls. Beads may also accumulate in subglacial cavities that develop as offshoots to the main tunnel. Where well-integrated esker networks carry large volumes of debris to the glacier margin they may link up with extensive ice-marginal depositional systems that include ice-contact deltas, subaqueous fans and MORAINES.

References

Shreve, R.L. (1972) Movement of water in glaciers, *Journal of Glaciology* 11, 205–214.

Warren, W.P. and Ashley, G.M. (1994) Origins of the ice-contact stratified ridges (eskers) of Ireland, *Journal of Sedimentary Research* A64, 433–449.

Further reading

Auton, C.A. (1992) Scottish Landform Examples – 6: The Flemington eskers, *Scottish Geographical Magazine* 108, 190–196.

Bannerjee, I. and McDonald, B.C. (1975) Nature of esker sedimentation, in A.V. Jopling and B.C. McDonald (eds) *Glaciofluvial and Glaciolacustrine Sedimentation*, 132–154, SEPM Special Publication 23.

Brennand, T.A. (1994) Macroforms, large bedforms and rhythmic sedimentary sequences in subglacial eskers, south-central Ontario: implications for esker genesis and meltwater regime, *Sedimentary Geology* 91, 9–55.

Evans, D.J.A. and Twigg, D.R. (2002) The active temperate glacial landsystem: a model based on Breiðamerkurjökull and Fjallsjökull, Iceland, *Quaternary Science Reviews* 21(20–22), 2,143–2,177.

Gorrell, G. and Shaw, J. (1991) Deposition in an esker, bead and fan complex, Lanark, Ontario, Canada, *Sedimentary Geology* 72, 285–314.

Price, R.J. (1969) Moraines, sandar, kames and eskers near Breiðamerkurjökull, Iceland, *Transactions of the Institute of British Geographers* 46, 17–43.

——(1973) *Glacial and Fluvioglacial Landforms*, Edinburgh: Oliver and Boyd.

Terwindt, J.H.J. and Augustinus, P.G.E.F. (1985) Lateral and longitudinal successions in sedimentary structures in the Middle Mause esker, Scotland, *Sedimentary Geology* 45, 161–188.

Thomas, G.S.P. and Montague, E. (1997) The morphology, stratigraphy and sedimentology of the Carstairs esker, Scotland, UK, *Quaternary Science Reviews* 16, 661–674.

DAVID J.A. EVANS

ESTUARY

Estuaries are unique ecosystems that provide spawning grounds for many organisms, feeding stops for migratory birds and natural filters to maintain water quality. Estuaries have value to humans for shipping and boating, settlements, erosion protection, recreation, mineral extraction and release of waste materials. Estuaries are generally considered areas where salt water from the ocean mixes with freshwater from land drainage but there are many definitions for the term (see Perillo 1995) reflecting the complex physical and biological processes present. Estuaries may be classified or described on the basis of numerous criteria, including entrance conditions (Cooper 2001), stage of development and degree of infilling (Roy 1984), morphology (Pritchard 1967; Fairbridge 1980), tidal range (Hayes 1975), vertical stratification and salinity structure (Cameron and Pritchard 1963). All these criteria affect the evolution of estuaries and the nature of transport of sediment and biota. The most common definition is Cameron and Pritchard (1963: 306) who define an estuary as 'a semi-enclosed coastal body of water having a free connection with the open sea and within which sea-water is measurably diluted with fresh water derived from land drainage'. The boundaries of an estuary can be defined by salinity (ranging from 0.1‰ at the head of the estuary and 30–35‰ at the mouth) or sedimentary facies and the processes that shape them. For example, Dalrymple *et al.* (1992) defined the upper boundary as the landward limit of tidal facies and the lower boundary as the seaward limit of marine facies.

From a geologic perspective, today's estuaries are recent features. Estuaries are a product of inherited factors (i.e. lithology) that influence the configuration of the estuarine basin and sediment type and availability; broad-scale controls such as climate and sea-level rise that influence rates of discharge and inundation; and, contemporary processes (wave, tide and river) that influence hydrodynamics and sediment transport. The position of estuaries is a result of fluctuations in sea-level rise, with sea-level elevation at or above current levels during interglacial periods and up to 150 m below present levels during glacial periods. Present-day estuaries are the result of sea-level rise and inundation of coastal lowlands following the last glacial period and the sea-level stillstand that began approximately 6,000 years ago. More recent regional sea-level histories have revealed both lowering and rising of sea level from stillstand levels.

Classification

Estuaries can be broadly classified as drowned river valleys, fjords, bar-built, and those formed by faulting or local subsidence (Pritchard 1967; Fairbridge 1980). Drowned river valley estuaries are found along the east coast of the United States (i.e. Delaware Bay and Chesapeake Bay) and in England (i.e. Thames and Mersey estuaries), France (i.e. Seine) and in Australia (i.e. Batesman Bay). Rivers eroded deep V-shaped valleys during the last glacial period that were subsequently inundated when melting ice sheets caused a rise in sea level. The planform and cross section of these estuaries are often triangular or funnel-shaped. In systems where sedimentation rates are less than rates of sea-level rise the river valley topography is maintained. Bar-built estuaries have a geologic history similar to drowned river valleys (the result of glacial incision and subsequent inundation by sea-level rise) but recent marine sediment transport (alongshore or cross shore) results in the creation of a barrier or spit across the mouth. The inlet at the mouth is small relative to the size of the shallow estuary created behind the barrier. In some cases the barrier may restrict exchange of water between the ocean and estuary except during high tides. Examples of this type of estuary can be found in the United States (i.e. Mobile Bay and Pamlico Sound) and in Australia (i.e. Clarence and Narooma Estuaries). Fjords are glacially-scoured U-shaped valleys that were subsequently inundated by a rising sea level. Most fjords possess a shallow rock sill near the mouth that forms an estuarine basin. Fjord-type estuaries are found in upper latitudes (i.e Oslo Fjord, Norway and Puget Sound, USA). Some estuaries are formed in valleys that were created by processes such as faulting (i.e. San Francisco Bay, USA) or subsidence.

Estuaries are located in micro-, meso- and macro-tidal environments. Planform morphology is an important control on the variation of tidal range and the magnitude of the tidal current within an estuary (Nichols and Biggs 1985). Convergence of the estuarine sides causes the tidal wave to compress laterally. In the absence of bed friction, the tidal range will increase. In the presence of friction, the tidal range will decrease. The relationship between convergence and friction control the amplitude of the tide within the estuary. In cases where convergence is greater than friction the tidal range and strength of the tide will increase toward the head of the estuary (hypersynchronous estuaries). In cases where convergence is less than friction the tidal range will decrease throughout the estuary (hyposynchronous estuaries).

Morphology

Wave, tide and river processes control the location of marine and river sediments in the estuary and the morphology of the sedimentary deposits. Conceptual models of estuarine morphology classify estuaries according to the relative contribution of these processes (see Dalrymple *et al.* 1992; Cooper, 1993) and are based, in part, on regional studies of estuarine sedimentation and morphology. These studies include tide-dominated,

macro-tidal estuaries (Dalrymple *et al.* 1990) and micro-tidal estuaries in wave-dominated (Roy 1984) and river-dominated (Cooper 1993) environments.

Tide-dominated estuaries are found in macro-tidal environments (tidal range > 4 m). They are generally funnel-shaped with wide mouths and high current velocities. Dalrymple *et al.* (1990) characterized the sedimentary characteristics of the macro-tidal Cobequid Bay–Salmon River Estuary, Canada. The axial sands are characterized by the presence of elongate tidal sand bars in the lower sector of the estuary that trend parallel to the dominant current direction. Sand flats and braided channels are located in the middle sector of the basin and a single channel is located in the river-dominated head of the estuary. Tidal currents are at a maximum in the inner part of the estuary. Sediments decrease in size from the mouth to the head. Dominant direction of sediment transport is landward with accumulation in the upper sector at the head of the estuary.

Wave-dominated estuaries are generally found in micro-tidal (tidal range < 2 m) environments (Roy 1984). In general, these estuaries have an upper sector near the head, where river processes, sediments and bedforms dominate, a lower sector near the mouth, where wave and tidal processes and marine sediments dominate, and a middle sector, where tidal currents dominate and both river and marine sediments are present. High wave and tidal energies at the mouth of an estuary can deposit sediment and restrict or completely prohibit exchange of water between the ocean and the estuary.

Mixed wave-tide estuaries (such as those in meso-tidal environments with a tidal range of 2–4 m) can be found behind barrier islands (Hayes 1975). The dominant sand bodies in meso-tidal estuaries are the deltas (ebb and flood) formed by tidal inlet processes. Within the estuary are meandering tidal channels and point bars and marsh deposits.

River-dominated estuaries do not display the characteristic downstream facies changes observed in wave- and tide-dominated estuaries, and energy levels may remain similar along the axis of the river valley (Cooper 1993). River-dominated estuaries can range from those completely dominated by river processes (river channels) to those that experience some marine inputs at the mouth.

Shoreline environments

Estuarine shoreline environments often occur in small isolated reaches with different orientations and with great variability in morphology, vegetation and rate of erosion. This variability results from regional differences in fetch characteristics, exposure to dominant and prevailing winds, variations in subsurface stratigraphy, irregular topography inherited from drainage systems, differential erosion of vegetation or clay, peat and marsh outcrops on the surface of the subtidal and intertidal zones, small-scale variations in submergence rates, effects of varying amounts of sediment in eroding formations and effects of obstacles to longshore sediment transport, such as headlands and coves, that define drift compartments (Nordstrom 1992). Differences in the gradient of wave energy between the low-energy (upper) and high-energy (lower) shorelines in an estuary and between the high-energy (windward) and low-energy (leeward) sides of an estuary also contribute to differences in the types of estuarine environments and their dimensions. Saltmarsh is likely to form on alluvium in the upper reaches of the estuary, on the upwind side of the estuary or on the downwind side of the estuary in the lee of headlands that provide protection from breaking waves. Beaches are likely to form on the downwind side of estuaries because there is sufficient energy in the locally generated waves to erode coastal formations or prevent plants from growing in the intertidal zone.

BEACHES may be unvegetated or partially vegetated and are composed of sand, gravel or shell. The dominant processes of sediment reworking on beaches in estuaries is usually locally generated waves, although refracted and diffracted ocean waves may be present. The best development of beaches occurs where relatively high wave energies have exposed abundant unconsolidated sand or gravel in the eroding coastal formations. Adequate source materials occur where these formations are moraine deposits, submerged glacial streams, coarse-grained fluvial deposits, and sand delivered by ocean waves and winds, such as the estuarine shorelines of spits and barrier islands. Beach formation is favoured where high ground protrudes into relatively deep water, where wave refraction and wave energy loss through dissipation on the bay bottom is minimal. Ocean waves that enter the estuary usually create beaches close to the inlets. Sediment transported into the estuary by ocean waves may form spits in the lee of headlands in the estuary. Beaches created by waves generated within estuaries are most common in shoreline re-entrants, where sediments can accumulate over time. Other beaches occur where sand is plentiful on the bayside of barriers enclosing the estuary, particularly on former recurves, subaerial overwash platforms and former oceanside dunes (Nordstrom 1992).

Beaches may form on the bay side of eroding marshes from coarse-grained sediment removed from the eroding substrate. Beaches may precede and favour marsh growth by creating spits that form low-energy environments landward of them. Both processes create a beach ridge shoreline that combines features characteristic of beach shorelines and marsh shorelines. Peat, representing the substrate of former marsh, is often exposed in outcrops on eroding beaches transgressing marshes. These outcrops are resistant because of the presence of fine-grained materials trapped by upward growth of the marsh and the binding effect of vegetation.

Dunes (see DUNE, COASTAL) form within estuaries only where beaches are sufficiently wide to provide a viable sediment source or where the shoreline is stable enough to allow ample time for slow accretion or to prevent wave erosion. Wave energy must be sufficient to prevent colonization of intertidal vegetation, but erosion cannot be too great for aeolian forms to survive. Onshore aeolian transport occurring between moderate-intensity storms may create only a thin aeolian cap on top of the backbeach or overwash platform.

Marshes are components of the intertidal profile affected by waves and tidal flows, and they bear many similarities with beaches, including the potential for cyclic exchange of sediment between the upper and lower parts of the profile and a tendency to buffer energy in a way that resists long-term morphological change (Pethick 1992). Marsh shorelines differ from beaches in that they are characterized by fine sediment sizes, low gradients and dissipative slopes. The occurrence of marshes, like beaches, depends on their environmental setting and mode of origin, defined by factors such as bedrock geology, availability of sediments and recent sea-level rise history (Wood *et al.* 1989). Examples of distinctive morphological marsh units determined by macro-scale differences within estuaries include fluvial marshes, occurring on the upper estuarine margins of rivers; bluff-toe marshes that form at the base of coastal bluffs; backbarrier marshes found behind barrier islands and spits; and transitional marshes where freshwater peatlands are colonized by saltmarsh (Wood *et al.* 1989).

Competition for human resource values of the estuarine shoreline has led to elimination of many natural environments. The conversion of some of the environments (especially bay bottoms and marshes) is now severely restricted by land use controls in many countries, but many estuaries are still threatened by human activities.

References

Cameron, W.M. and Pritchard, D.W. (1963) Estuaries, in M.N. Hill (ed.) *The Sea*, 306–324, New York: Wiley Interscience.

Cooper, J.A.G. (1993) Sedimentation in a river dominated estuary, *Sedimentology* 40, 979–1,017.

——(2001) Geomorphological variability among microtidal estuaries from the wave-dominated South African coast, *Geomorphology* 40, 99–122.

Dalrymple, R.W., Knight, R.J., Zaitlin, B.A. and Middleton, G.V. (1990) Dynamics and facies model of a macrotidal sand-bar complex, Cobequid Bay–Salmon River estuary (Bay of Fundy), *Sedimentology* 37, 577–612.

Dalrymple, R.W., Zaitlin, B.R. and Boyd, R. (1992) Estuarine facies models: conceptual basis and stratigraphic implications, *Journal of Sedimentary Petrology* 62, 1,130–1,146.

Fairbridge, R.W. (1980) The estuary: its definition and geodynamic cycle, in E. Olausson and I. Cato (eds) *Chemistry and Biogeochemistry of Estuaries*, 1–35, New York: Wiley.

Hayes, M.O. (1975) Morphology of sand accumulation in estuaries: an introduction to the symposium, in L.E. Cronin (ed.) *Estuarine Research, Vol. II*, 3–22, New York: Academic.

Nichols, M.M. and Biggs, R.B. (1985) Estuaries, in R.A. Davis (ed.) *Coastal Sedimentary Environments*, 77–186, New York: Springer-Verlag.

Nordstrom, K.F. (1992) *Estuarine Beaches*, London: Elsevier.

Perillo, G.M.E. (ed.) (1995) *Geomorphology and Sedimentology of Estuaries*, New York: Elsevier.

Pethick, J.S. (1992) Saltmarsh geomorphology, in J.R.L. Allen and K. Pye (eds) *Saltmarshes: Morphodynamics, Conservation and Engineering Significance*, 41–62, Cambridge: Cambridge University Press.

Pritchard, D.W. (1967) What is an estuary: physical viewpoint, in G.H. Lauff (ed.) *Estuaries*, 3–5, Washington, DC: American Association for the Advancement of Science.

Roy, P.S. (1984) New South Wales estuaries: their origin and evolution, in B.G. Thom (ed.) *Coastal Geomorphology in Australia*, 99–121, New York: Academic.

Wood, M.E., Kelley, J.T. and Belknap, D.F. (1989) Patterns of sediment accumulation in the tidal marshes of Maine, *Estuaries* 12, 237–246.

N.L. JACKSON

ETCHING, ETCHPLAIN AND ETCHPLANATION

The word 'etching' generally means corroding a surface by aggressive reagents and is used in geomorphology to describe progressive rock decomposition which occurs within deep weathering profiles. In particular, it is applied to situations where rocks differ in their resistance to chemical decay and consequently thickness of a weathering mantle is highly variable over short distances. Removal of products of deep weathering will expose bedrock surface, the topography of which is the direct result of etching, thus it is an 'etched surface'. At an early stage of development of geomorphology, when focus on planation surfaces and peneplains was pre-eminent, etched surfaces were visualized as surfaces of rather low relief and thought of as a special category of a PENEPLAIN, produced by deep-reaching rock decay followed by stripping of weathering mantle. For surfaces of this origin, the term 'etchplain' was proposed by B. Willis and E.J. Wayland, working in East Africa in the 1930s. Accordingly, the process of producing an etchplain through weathering and stripping has later become known as 'etchplanation'.

The impact of the concept of etching and etchplanation on general geomorphology was initially limited, mainly due to the association of etchplains with peneplains, remoteness of original study areas, and restricted access to early publications. Furthermore, no applications for extratropical areas were offered and etchplains were considered as of local importance and specific for low latitudes. Proposals to restrict the usage of the term 'etchplain' to areas of exposed rock, i.e. completely stripped of weathering products, added to the slow progress and general downplaying of the significance of etchplains in geomorphology.

The situation began to change with the arrival of the paper by Büdel (1957) which is noteworthy for a number of reasons, although the very term 'etchplain' was not used. First, Büdel made clear that he was applying the weathering/stripping concept to entire landscapes rather than to limited areas within them or to individual landforms. Second, he suggested that many upland surfaces in middle and high latitudes are inherited Tertiary etchplains, and hence extended the applicability of the concept outside the tropics. Third, he pointed out that transition from the phase of dominant weathering to the phase of dominant stripping might be associated with major environmental changes, whose profound impact for landform development was realized only later. Fourth, it contributed to the appreciation of the concept by the Central European geomorphological community, which was reflected in numerous detailed studies shortly after.

Realization of the crucial role of deep weathering and SAPROLITE development in shaping most tropical landscapes, achieved in the 1960s, led to the expansion of the original ideas of Wayland and Willis, so different types of landscapes could be described. The proposed classification, in the form subsequently modified by its author himself (Thomas 1989), include:

- *Mantled etchplain*: weathering mantle is ubiquitous and virtually no bedrock is exposed. Weathering progressively attacks solid rock at the base of the mantle, moulding the etched surface which is to be exposed later, but the mantle can also be relict.
- *Partly stripped etchplain*: develops from mantled etchplain through selective removal of the weathering mantle and exposure of bedrock surface, but part of the original saprolite remains. The proportion of areas still underlain by saprolite may vary from 10 to almost 100 per cent.
- *Stripped etchplain*: most of the bedrock is exposed from beneath a weathering mantle and only isolated patches of saprolite are left (< 10 per cent of the area). These characteristics conform with the original definition by Wayland.
- *Complex etchplain*: includes a few variants, in which deeply incised valleys may be present (incised etchplain), or removal of saprolite is accomplished by pedimentation (pedimented etchplain), or a new generation of weathering mantles begins to form (re-weathered etchplain).
- *Buried etchplain*: one which has been covered by younger sediments or lava flows.
- *Exhumed etchplain*: one which has been re-exposed after burial.

One important terminological problem has been noted, that a stripped surface is rarely a plain but tends to show some relief, which reflects differential progress of etching (see Figure 59). This happens in particular, if bedrock is lithologically varied or various structures, e.g. fractures, are differentially exploited by weathering. In many granite areas, stripped surfaces are typified by domes, TORS, basins and boulder piles, and to call them 'etchplains' would be inappropriate and misleading. Therefore the term 'etchsurface' is recommended for use wherever evacuation of weathering mantles reveals a varied topography.

Etchplanation, and in particular the transition from the weathering phase to the stripping phase, is commonly linked with major external changes experienced by a landscape, related either to a change in tectonic regime or to environmental changes. It is envisaged that mantled etchplains form and exist during long periods (up to 10^9 yr) of stability, whereas stripping is initiated by uplift, or climatic change towards drier conditions, and is accomplished over much shorter timescales (10^5–10^7 yr). In this view, major external disturbances are essential for the formation of etchplains and static nature of planate landsurfaces might be implied. This position is contradicted by field evidence of geomorphic activity, hence an idea of 'dynamic etchplanation' has been introduced to emphasize ongoing landscape development through etching and stripping (Thomas and Thorp 1985). Key points made are simultaneous weathering and removal of its products, lowering of both interfluves and valley floors, continuous sediment transfer, redistribution and temporal storage of weathering products, and the importance of minor environmental disturbances.

From the 1980s onwards, following the progress in weathering studies, the etching concept has been extended away from tropical plains to the much wider range of settings. Emphasis on the process of deep weathering rather than on the ultimate form of a plain has made it possible to see geomorphic development of many low-latitude mountain ranges of moderate relief as being accomplished by differential etching. Deep weathering is facilitated by strong groundwater movement, steep hydraulic gradient, tensional fracture patterns and numerous lines of weakness within bedrock, whereas landslides play a major part in removal of the saprolite. Realization that formation of thick sandy weathering mantles (see GRUS) can effectively take place outside the tropics has opened the way to interpret several middle to high latitude terrains, with no or extremely remote history of tropical conditions, as etchsurfaces or etchplains, even if the very term has not always been used (Pavich 1989; Lidmar-Bergström 1995; Migoń and Lidmar-Bergström 2001).

Over the years, the idea of etching and etchplanation has evolved from being a mere specific, 'tropical' variant of peneplanation to the status of an autonomic concept, capable of accounting for both planate and topographically complex landsurfaces, integrating tectonic and climatic controls, linking historical and process geomorphology. Despite what the name and early history suggest, it must not be regarded as focusing on explaining the origin of planation surfaces. Nor does it compete with other planation theories, as for instance pedimentation may be a means of stripping. To the contrary, long-term etching and stripping may, and in many places do, lead to the differentiation and increase of relief. Depending on local lithology, tectonic setting and environmental history, long-term etching may transform an initial landscape into a range of topographies, from plains to even mountainous. Therefore, it is the evidence of former or present deep weathering which is a prerequisite for a terrain to be identified as an etchsurface, and not any particular assemblage of landforms.

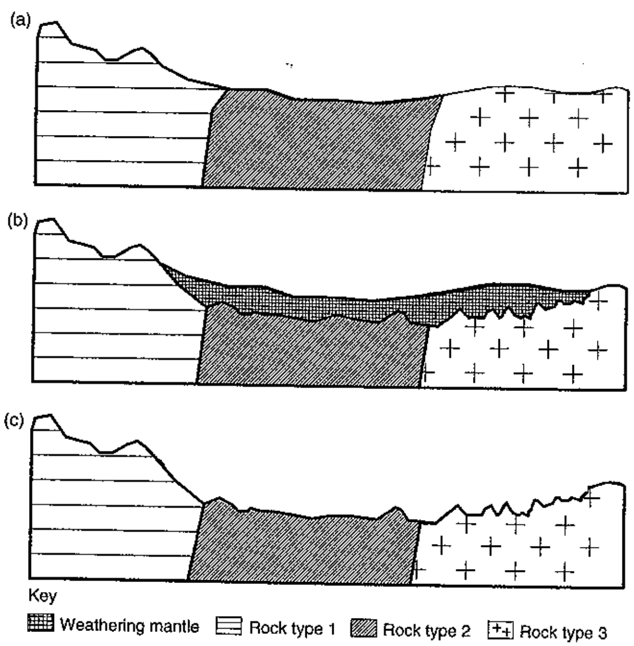


Figure 59 Depending on (a) bedrock characteristics and their susceptibility to (b) selective deep weathering, etching may produce surfaces of (c) various types, for instance inselberg-dotted plains (middle) or multi-convex, hilly areas (right)

References

Büdel, J. (1957) Die 'doppelten Einebnungsflächen' in den feuchten Tropen, *Zeitschrift für Geomorphologie N.F.* 1, 201–1, 228.

Lidmar-Bergström, K. (1995) Relief and saprolites through time on the Baltic Shield, *Geomorphology* 12, 45–61.

Migoń, P. and Lidmar-Bergström, K. (2001) Weathering mantles and their significance for geomorphological evolution of central and northern Europe since the Mesozoic, *Earth Science Reviews* 56, 285–324.

Pavich, M.J. (1989) Regolith residence time and the concept of surface age of the Piedmont 'peneplain', *Geomorphology* 2, 181–196.

Thomas, M.F. (1989) The role of etch processes in landform development, *Zeitschrift für Geomorphologie N.F.* 33, 129–142 and 257–274.

Thomas, M.F. and Thorp, M. (1985) Environmental change and episodic etchplanation in the humid tropics: the Koidu etchplain, in I. Douglas and T. Spencer (eds) *Environmental Change and Tropical Geomorphology*, 239–267, London: George Allen and Unwin.

Further reading

Adams, G. (ed.) (1975) *Planation Surfaces* Benchmark Papers in Geology, 22.

Bremer, H. (1993) Etchplanation, Review and Comments of Büdel's model, *Zeitschrift für Geomorphologie N.F. Supplementband* 92, 189–200.

Thomas, M.F. (1994) *Geomorphology in the Tropics*, Chichester: Wiley.

Twidale, C.R. (2002) The two-stage concept of landform and landscape development involving etching: origin, development and implications of an idea, *Earth Science Reviews* 57, 37–74.

SEE ALSO: granite geomorphology; inselberg; planation surface; tropical geomorphology

PIOTR MIGOŃ

EUSTASY

The concept of 'eustatic' changes in sea level, implying vertical displacements of the ocean surface occurring uniformly throughout the world, was introduced by Suess (1885–1909). Global changes in sea level depend in fact on a combination of factors (changes in the quantity of oceanic water, deformation of the shape of the oceanic basin, variations in water density, and dynamic changes affecting water masses) which operate globally, regionally or locally on different timescales.

The quantity of ocean water is controlled mainly by climate, that may cause the development or melting of huge continental ice sheets. According to IPCC (2001), the present volume of continental ice can be estimated at about 29 million cubic kilometres, equivalent to *c.*70 metres water depth in the oceans. At the time of the last glacial maximum, some 20 ka ago, when the global sea level was estimated about 120 metres lower than now, the continental-ice volume must have been more than double the present.

Daly (1934) stressed the importance of changes in sea level and of glaci-isostatic effects (see ISOSTASY) accompanying the last deglaciation phase, with uplift in areas of ice melting and subsidence in a wide peripheral belt. During the last decades, improved global isostatic models based on ice volumes and water depths have been developed (e.g. Lambeck 1993; Peltier 1994), demonstrating that ice-volume changes imply vertical deformation of Earth crust which is highly variable regionally.

Mörner (1976) supported anew the old notion of geoidal changes, suggesting that displacements of the bumps and depressions revealed by satellites on the ocean surface topography could cause differences between coastal areas in the relative sea-level history.

Recently, the analysis of satellite observations, especially by Topex/Poseidon, have shown that the level of the ocean surface can be closely correlated with sea-surface temperature. The resulting sea-surface topography is highly variable, with sea-level rise in certain areas, and sea-level fall in other areas. Steric effects, which depend on the temperature (and density) of the whole water column are also highly variable. Analysis of the dynamic behaviour of water masses and their displacements would bring similar results.

SEA-LEVEL changes are therefore not uniform, but variable over several temporal and spatial scales. Sea level may therefore change from place to place in the ocean, and even more in coastal areas, where hydro-isostatic movements are controlled by the water depth on the continental shelf. In short, there is now general agreement that no coastal area exists where the local sea-level history could be representative of the global eustatic situation. Worldwide or simultaneous sea-level changes, therefore, do not exist. They are an abstraction.

In spite of this field evidence, the concept of eustasy is not an obsolete one, because the estimation of global sea-level changes, even if obtained with some approximation, may have many useful applications in geosciences, in relation to climate, tectonics, paleo-environmental, and also near-future environmental changes. If eustatic variations cannot be specified from coastal field data, the estimation of the changes in the quantity of ocean water is possible using geochemical analysis of marine sediments. The $\delta^{18}O$ content in fossil foraminifera shells cored from the deep ocean floor depends on the salinity and temperature of sea water at the time they lived. If benthic species are selected, temperature changes will be minimized and $\delta^{18}O$ will depend mainly on salinity, i.e. on the quantity of fresh water held up in continental ice sheets. Such calculation makes the estimation of approximate eustatic changes possible, with assumptions, and with an accuracy that depends on the resolution of geochemical measurements, i.e. with an uncertainty range, for sea level, of the order of ± 10 m. Such precision may seem relatively poor, if compared to what can be obtained at a local scale from the study of former shorelines data. In addition, tectonic, isostatic, steric and hydrodynamic factors are neglected. Nevertheless, continuous oceanic cores have the great advantage that they can cover long time sequences, making approximate estimations of eustatic oscillations possible for periods even longer than the whole Quaternary. According to Milankovitch's astronomical theory, major climatic changes have an astronomical origin, with cycles of near 100 ka for orbital eccentricity, 41 ka for orbital obliquity, and 23 ka and 19 ka for precession phenomena. The age of the climate oscillations deduced from oceanic cores is generally estimated with good accuracy, through calibration with selected astronomical (e.g. insolation at 65°N) curves.

Even with some approximation, eustatic oscillations can be very useful to coastal geomorphologists, e.g. to those who study sequences of datable raised marine terraces in uplifting areas. Each terrace, especially if made of coral reefs, can be considered to have developed when the rising sea level overtook the rising land, and therefore corresponds to a sea-level transgression peak. In this manner, the sea level relative to a stable oceanic floor can be extracted from each dated section if the uplift rate for that section is known (Chappell and Shackleton 1986).

Estimates of eustatic changes during the last century have been attempted by many authors, mainly using tide-gauge records. Discrepancies arising from different analysis methods (for a critical review, see Pirazzoli 1993) remain high enough, however, for IPCC (2001) not to choose between a recent sea-level rise of $1.0\,mm\,yr^{-1}$, or an upper bound of $2.0\,mm\,yr^{-1}$, or a central value of $0.7\,mm\,yr^{-1}$ estimated independently from observations and models of sea-level rise components.

Satellite data are more reliable than tide gauges for global calculations. According to Topex/Poseidon, a global sea-level rise of $2.5 \pm 0.2\,mm\,yr^{-1}$ can be inferred between January 1993 and December 2000 (Cabanes *et al.* 2001). However, El Niño events produce significant oscillations in the global sea level trend and a few decades of additional records would be necessary before a reliable assessment of the present-day eustatic trend can be made with some confidence.

For the next century, eustatic predictions are based on climatic models and scenarios of greenhouse gas emissions. The most recent estimate (IPCC 2001) is of a global sea-level rise of 0.09 to 0.88 m for the period from 1990 to 2100, with a central value of 0.48 m. This estimate includes the variation in the quantity of oceanic water and steric effects, but is exclusive of vertical land motion and hydrodynamic effects. Therefore, it may have very little to do with the relative sea level experienced on a regional basis or at single sites.

References

Cabanes, C., Cazenave, A. and Le Provost, C. (2001) Sea level rise during past 40 years determined from satellite and in situ observations, *Science* 294, 840–842.

Chappell, J. and Shackleton, N.J. (1986) Oxygen isotopes and sea level, *Nature* 324, 137–140.

Daly, A. (1934) *The Changing World of the Ice Age*, New Haven: Yale University Press.

IPCC (2001) *Climate Change 2001: The Scientific Basis*, Cambridge (UK) and New York: Cambridge University Press.

Lambeck, K. (1993) Glacial rebound and sea-level change: an example of a relationship between mantle and surface processes, *Tectonophysics* 223, 15–37.

Mörner, N.A. (1976) Eustasy and geoid changes, *Journal of Geology* 84(2), 123–151.

Peltier, W.R. (1994) Ice age paleotopography, *Science* 265, 195–201.

Pirazzoli, P.A. (1993) Global sea-level changes and their measurement, *Global and Planetary Change* 8, 135–148.

Suess, E. (1885–1909) *Das Antlitz der Erde, 3 vols*, Wien.

Further reading

Pirazzoli, P.A. (1996) *Sea-Level Changes – The Last 20,000 Years*, Chichester: Wiley.

P.A. PIRAZZOLI

EVORSION

The erosion of rock or sediments in a river or streambed, by the impact of clear water carrying no suspended load. The process of evorsion often results in the formation of pot-holes (evorsion pot-holes) within the streambed, due to the action of eddies and vortices. The predominant processes involved in evorsion are hydraulic action and fluid stressing.

Further reading

Aengeby, O. (1952) Recent, subglacial and laterglacial pothole erosion (evorsion), *Lund Studies in Geography, Series A, Physical Geography* 3, 14–24.

STEVE WARD

EXFOLIATION

The shedding of material in scales or layers, it is often used interchangeably with sheeting or onion-skin weathering. Exfoliation of rock has been attributed to various causes including UNLOADING, insolation and HYDRATION (see INSOLATION WEATHERING). It is a process that has some applied significance and is, for example, a consideration in road, tunnel and dam construction where excavation can cause pressure release and fracturing to occur (Bahat *et al.* 1999). Exfoliation occurs at a variety of spatial scales from thin (<cm) scaling from boulders to mega-form some metres in size (Bradley 1963).

Plate 42 As a consequence of pressure release resulting from the erosion of overlying material, this granite near Kyle in Zimbabwe is being broken up into a series of curved sheets which parallel the land surface

References

Bahat, D., Grossenbacher, K. and Karasaki, K. (1999) Mechanism of exfoliation joint formation in granitic rocks, Yosemite National Park, *Journal of Structural Geology* 21, 85–96.

Bradley, W.C. (1963) Large-scale exfoliation in massive sandstones of the Colorado Plateau, *Geological Society of America Bulletin* 74, 519–528.

A.S. GOUDIE

EXHUMED LANDFORM

Landforms that have been covered by sedimentary strata or volcanic rocks and then re-exposed are called exhumed. Exhumed landforms of different age are common within Precambrian shields. Oldest exhumed landforms in Australia are encountered below Proterozoic covers. Flat surfaces are often exhumed from below Lower Palaeozoic rocks on the Laurentian and Baltic shields, while etched (deeply weathered) more or less hilly surfaces extend from below Jurassic or Cretaceous rocks in Minnesota, USA, along parts of the Greenland west coast and in southern Sweden. Hilly relief is exhumed from Neogene sediments in south Poland. Glacially polished surfaces extend from below Upper Precambrian strata in northern Norway, Ordovician strata in the Sahara, and Permian strata on the Gondwana continents. Palaeokarstic features are exhumed from below Carboniferous strata in eastern Canada and in south Germany they make up the Kuppenalb, exhumed from a Cretaceous cover. Exhumed landforms give important information on past processes and their recognition is necessary for correct interpretation of present landscapes. Exhumed denudation surfaces are also important geomorphic markers for studies of Cainozoic uplift and erosion.

Further reading

Ambrose, J.W. (1964) Exhumed palaeoplains of the Precambrian shield of North America, *American Journal of Science* 262, 817–857.

Fairbridge, R.W. and Finkl, C.W. (1980) Cratonic erosional unconformities and peneplains, *Journal of Geology* 88, 69–86.

Lidmar-Bergström, K. (1996) Long term morphotectonic evolution in Sweden, *Geomorphology* 16(1), 33–59.

Migon, P. (1999) Inherited landscapes of the Sudetic Foreland (SW Poland) and implications for reconstructing uplift and erosional histories of upland terrains in Central Europe, in B.J. Smith, W.B. Whalley and P.A. Warke (eds) *Uplift, Erosion and Stability: Perspective on Longterm Landscape Development*, Geological Society London Special Publications, 162, 93–107.

Peulvast, J.-P., Bouchard, M., Jolicoeur, G. and Schroeder, J. (1996) Palaeolandforms and morphotectonic evolution around the Baie de Chaleurs (eastern Canada), *Geomorphology* 16(1), 5–32.

KARNA LIDMAR-BERGSTRÖM

EXPANSIVE SOIL

Most clay soils experience a volume change on wetting and drying. The soils with the relatively inactive clay minerals, such as kaolinite, only produce a modest volume change; but soils containing montmorillonite and other smectite minerals, can have considerable changes of volume; expanding on wetting and shrinking on drying. This causes widespread construction problems because of damage to buildings and other structures, but also accounts for certain geomorphological features such as GILGAI.

The clay particles in soils carry electrical charges, and this accounts for their interesting relationship with water. Clay mineral particles tend to be negatively charged and to attract the cations in the soil water. These cations are hydrated because the negative end of the polarized water molecule is attracted to the charged ion. So, via the activity of the cations, the clay minerals attract water, and this confers on the clay systems the property of plasticity. This is measured via the plasticity index PI, which is low, perhaps around 20, for the inactive kaolinites, illites and chlorites, but high, perhaps up to 200, for the montmorillonites. It is the high PI systems which dominate in expansive soils. The structure of montmorillonite is such that water enters between the clay layers and generates considerable swelling force. The uplift pressure in undisturbed montmorillonite clays can be up to 0.1–0.6 MN m−2. These expansive pressures easily exceed the loads imposed by small structures such as single-family houses and single-storey schools. It is damage to these small structures which generates the vast costs caused by expansive soil effects. In the USA costs of over $2 billion a year are cited. This is about twice the cost of flood or landslide damage, and more than twenty times the cost of earthquakes.

In the USDA Soil Taxonomy system of soil classification the expansive soils fall into the order Vertisols, and they are defined as cracking soils; mineral soils that have been strongly affected by argillipedoturbation, i.e. mixing by the shrinking and swelling of clays. This normally requires alternate wetting and drying in the presence of >30 per cent clay, much of which, typically, is montmorillonite or some other smectite mineral. If not irrigated, these soils have cracks at least 1 cm wide at a 50 cm depth at some season in most years. Vertisols form the Black Cotton soils of north-west India; they form from the basalts of the Deccan plateau, under the influence of tropical weathering. These soils are classified as Usterts, i.e. ustic (dryish) vertisols; and so are the soils in east Australia which comprise the other large occurrence.

Regions underlain extensively by expansive clays can often be recognized by a distinctive

microtopography called gilgai. Where undisturbed by humans, gilgai can be easily distinguished on aerial photography either as an irregular network of microridges, or, where the slope is greater than 1 per cent, as a pattern of downslope linear ridges and troughs. In Australia gilgai relief has been observed to reach over 3 metres. Gilgai microrelief can be used as a rapid means of mapping regions where a significant potential hazard from clay soil expansion can be expected.

The potential volume change of soils is controlled by a number of factors: (1) the type of clay, amount of clay, cations present and clay particle size, (2) density; dense or consolidated soils swell the most, (3) moisture content; dry soils swell more than moist soils, (4) soil structure; remoulded soils swell more than undisturbed soils, (5) loading; schools and houses with lightly loaded foundations are the most susceptible.

There are a variety of tests for expansive soils but one of the most reliable is the oedometer (consolidometer) test. In this test compacted soils are loaded and then wetted and the expansive pressures produced are measured. A simple classification can be produced:

$0–0.15\,MN\,m^{-2}$	= non-critical
$0.15–0.17\,MN\,m^{-2}$	= marginal
$0.17–0.25\,MN\,m^{-2}$	= critical
$>0.25\,MN\,m^{-2}$	= very critical

It is possible to recognize expansive soils in the field; some factors to look out for are:

Under dry soil conditions

- Soil hard and rocklike; difficult/impossible to crush by hand
- Glazed, almost shiny surface where previously cut by scrapers, digger teeth or shovels
- Very difficult to penetrate with auger or shovel
- Ground surface displays cracks occurring in a more or less regular pattern
- Surface irregularities cannot be obliterated by foot pressure

Under wet soil conditions

- Soil very sticky; exposed soil will build up on shoe soles
- Can be easily moulded into a ball by hand; hand moulding will leave a nearly invisible residue on the hands after they dry
- A shovel will penetrate soil quite easily and the cut surface will be very smooth and shiny
- Freshly machine scraped or cut areas will tend to be very smooth and shiny
- Heavy construction equipment will develop a thick soil coating that may impair their function.

These high PI, high montmorillonite soils can be stabilized by lime addition. This causes cation replacement and the soils become more rigid.

Further reading

Chen, F.H. (1988) *Foundations on Expansive Soils*, Amsterdam: Elsevier.

Fanning, D.S. and Fanning, M.C.B. (1989) *Soil: Morphology, Genesis, and Classification*, New York: Wiley.

Proceedings of the 7th International Conference on Expansive Soils, Dallas, Texas, (1992) 2 vols.

Yanagisawa, E., Moroto, N. and Mitachi, T. (eds) (1998) *Problematic Soils*, section on expansive and collapsible soils, 253–384, Rotterdam: Balkema.

IAN SMALLEY

EXPERIMENTAL GEOMORPHOLOGY

Experimental geomorphology is the study, under experimental conditions, of a representation of a selected geomorphological feature or landscape. The term 'representation' is intended to cover full-scale features, scale models (hardware models), and numerical constructs. This definition raises the question 'what constitutes a geomorphological experiment?' Writing in a geomorphological context, Church (1984: 563) defined a scientific experiment as 'an operation designed to discover some principle or natural effect, or to establish or controvert it once discovered'. This activity differs from casual observation in that the phenomena observed are, to a critical degree, controlled by human agency, and from systematically structured observations in that the results must bear on the verity of some conceptual generalization about the phenomena. That definition leads to specific criteria for an experiment:

1. There must be a conceptual model of the processes or relations of interest that will ultimately be supported or refuted by the experiment, giving rise to:
2. Specific hypotheses about landforms or landforming processes that will be established or falsified by the experiment. (If the conceptual

model is a well-developed theory, the hypotheses will constitute exact predictions.)

To test the hypotheses, three further conditions are required:

3 Definitions must be given of explicit geomorphological properties of interest and operational statements of the measurements that will be made on them (sufficiently completely that replicate measurements might be made elsewhere);
4 A formalized schedule must be established of measurements to be made in conditions that are controlled insofar as possible to ensure that the remaining variability be predictable under the research hypotheses;
5 A scheme must be specified for analysis of the measurements that will discriminate the possibilities in (2).

The critical condition is the fourth one. Under this rubric, Church recognized two types of geomorphological experiment:

(a) Intentional, controlled interference with the natural conditions of the landscape in order to obtain unequivocal results about a limited set of processes that change the landscape;
(b) Statistically structured replication of observations in the landscape so that extraneous variability is effectively controlled or averaged over the experimental units.

Experiments of the second type entrain the capacity of statistical experimental design to discriminate and classify information in contexts where variability cannot actively be controlled. Ecologists face problems similar to those posed by geomorphology and have developed a sophisticated understanding of statistical experimentation (e.g. Hairston 1989) in order to deal with them.

The space and time scales associated with the development of most landforms effectively exclude them from experimental study. For this reason a definition of experimental geomorphology given by Schumm *et al.* (1987: 3) includes a careful exemption from strict experimental accountability. They wrote that experimental geomorphology is 'study, under closely monitored or controlled experimental conditions', accepting close observation as sufficient to establish a geomorphological experiment. Slaymaker (1991) similarly accepted 'formally structured', though not actively controlled, field studies as satisfactorily experimental exercises; exercises characterized by Church rather as formal case studies.

Schumm *et al.* (1987) also strongly implied that the objects of experimental manipulation would normally be small, or deliberately reduced scale, examples of field landforms. Experimental control is much more easily arranged in such cases, and their major summary book is entirely preoccupied with model studies. Models represent satisfactory experimental tests of hypotheses designed to explain features of the full-scale landscape provided that formal scaling criteria establish that the results can faithfully be extrapolated. Schumm *et al.* indicated no such requirement, even though there is a long history of scale model investigations in Earth science. Instead, they proposed two other perspectives. They suggested that reduced scale landforms be regarded simply as small prototypes. This does not release the investigator from formal scale constraints for extrapolation. They also explored the concept of model studies as analogues of full-scale systems or as studies in which there remains 'similarity-of-process' (after Hooke 1968, who elided the two perspectives). They argued that the model results might be extrapolated at least qualitatively to increase understanding of the full-scale landscape. Yet they recognized that changes in physical processes that undermine the supposed similarity may occur over large changes in scale. The approach, whilst it may be fruitful of ideas, suffers from inability to achieve exact predictions, or even to confirm unequivocal similarity of process, which disqualifies it as an experimental approach under the criteria given above.

There is, in fact, a tolerably well-developed body of conceptual and empirical studies of scale effects in geomorphology and hydrology (Church and Mark 1980), whilst formal scaling criteria have been investigated for hydrological and sedimentation processes at hillslope, channel and catchment scale. Formal scaling of generic model results (a generic model is one that captures the essential elements of a prototype whilst not conforming in inessential details with any particular prototypical example) ought to be possible in many problems.

In the field, one immediately faces the critical question whether experimental control can be adequately established. Geomorphological processes are driven by weather, which cannot feasibly be controlled at scales beyond that of a plot of order $10^2\,m^2$ (which might be enclosed). But variable forcing by weather is a fundamental feature of geomorphological systems, so one that is driven by artificially controlled weather is in some sense an unrealistic environmental system. Active manipulation should perhaps rather be focused on characteristics of the landform or landscape. Then it will usually be essential to establish a parallel reference or control case in which no manipulation is undertaken, in order that the effects of manipulation of the experimental system may be separated from the effects of variable weather.

It is helpful to differentiate landform and landscape studies. The former present more tractable space and time scales, and are more apt to represent environmental systems sufficiently simple to be amenable to control. At relatively small scales, successful experiments include plot studies of soil erosion, ground surface manipulations to investigate periglacial processes, and local applications of artificial precipitation or drainage adjustments to study effects on erosion or slope stability.

The centre of interest in geomorphology lies in transformations of landscape, which may be addressed through catchment experiments. There is far more experience with them than with any other full-scale experimental arrangement in Earth science (see Rodda 1976, for a historical perspective and critique). Much of the difficulty associated with the use of experimental catchments lies in establishing similarity between a treated catchment and its control, and in deciding how far observed results may be extended to the rest of the landscape. At the base of both issues lies the problem of establishing or measuring similarity between landscapes (Church 2003). Despite the known difficulties, the recognition of small catchments as the fundamental unit for most geomorphological process investigations ensures continuing effort to establish experimentally rigorous investigations at this scale.

Geomorphological experiments may be established inadvertantly. It is important not to overlook the potential value of landform or landscape manipulations undertaken for other purposes that nevertheless can be interpreted satisfactorily in terms of the requirements for an experiment. By this means, far larger systems than might ever be available for deliberate experimental manipulation may be studied. Examples include river rectification, certain water regulation projects, and certain changes in land surface condition. In such cases, it is important to establish a satisfactory reference comparison. Sometimes, this might be a before/after comparison in the same system; otherwise, a parallel reference case must be identified.

Numerical experimentation – the construction and operation of numerical models of geomorphological processes – represents a means by which complete control can indeed be gained over the conditions that drive landscape development. The penalty, of course, is uncertainty whether the numerical model faithfully represents all of the significant processes at work in the world. There also remain significant questions surrounding the means by which model outcomes may be compared with real landscapes, similar to those encountered in comparisons between real landscapes. Nevertheless, numerical modelling holds the promise to be an effective means to establish experimental control in geomorphological studies, especially ones concerning the development of entire landscapes over geomorphologically significant periods of time.

References

Church, M. (1984) On experimental method in geomorphology, in T.P. Burt and D.E. Walling (eds) *Catchment Experiments in Fluvial Geomorphology*, 563–580, Norwich: Geo Books.
——(2003) What is a geomorphological prediction?, in P.R. Wilcock and R.L. Iverson (eds) *Prediction in Geomorphology*, American Geophysical Union, Geophysical Monographs series.
Church, M. and Mark, D.M. (1980) On size and scale in geomorphology, *Progress in Physical Geography* 4, 342–390.
Hairston, N.G. Sr. (1989) *Ecological Experiments*, Cambridge: Cambridge University Press.
Hooke, R.L. (1968) Model geology: prototype and laboratory streams. Discussion, *Geological Society of America Bulletin* 79, 391–394.
Rodda, J.C. (1976) Basin studies, in J.C. Rodda (ed.) *Facets of Hydrology*, 257–297, London: Wiley-Interscience.
Schumm, S.A., Mosley, M.P. and Weaver, W.E. (1987) *Experimental Fluvial Geomorphology*, New York: Wiley-Interscience.
Slaymaker, O. (1991) The nature of geomorphic field experiments, in O. Slaymaker (ed.) *Field Experiments and Measurement Programs in Geomorphology*, 7–16, Rotterdam: Balkema.

MICHAEL CHURCH

EXTRATERRESTRIAL GEOMORPHOLOGY

The term 'extraterrestrial geomorphology' was not included in the 1968 *Encyclopedia of Geomorphology*. Indeed, at first reading, this term might seem to be an oxymoron. Should not a science of Earth forms (geomorphology) exclude those forms that are beyond the Earth (extraterrestrial)? The answer to this question depends on how one views the nature of science. Is a science more about methods and attitudes of study, or is it more about the organized accumulation of facts associated with specific subject matter? While organized fact accumulation might require precise definition as to the location of its subject matter, the methods and attitudes of geomorphology are readily conceived as extending to landforms on Earth-like planets (Baker 1993), if only better to understand Earth's landforms. To the extent that geomorphology emphasizes methods and attitudes for the study of landforms and landscapes, then it is no more restricted to studying Earth's landforms and landscapes than geometry is restricted to measuring the Earth's mathematical form.

Despite the immense excitement of planetary exploration during the 1960s, 1970s, and 1980s, there was a conspicuous lack of attention to planetary surfaces by mainstream geomorphologists. Nearly all the study of newly discovered landscapes was performed by scientists with very little background in geomorphology. More recently, increased attention to extraterrestrial geomorphology is indicated by Dorn's (2002) analysis of citations to late twentieth-century research works in geomorphology. Two of the 'top ten' cited geomorphology papers in recent years were directly concerned with topics in extraterrestrial geomorphology.

Historical and philosophical perspectives

It was not long after Galileo Galilei (1564–1642) first used a telescope to observe curious circular depressions on the Moon that Robert Hooke performed the first known geomorphological experiments to explain the origin of those depressions. Hooke was an intellectual adversary of Sir Isaac Newton, and, unlike Newton, he had a great interest and considerable talent for geology and geomorphology. In 1665 he compared the newly discovered lunar craters to (1) the cooled surface crust of melted gypsum, which was disrupted by bursting bubbles, and (2) the impact of musket balls and mud pellets into a clay-water target material. Using analogy as his mode of reasoning, Hooke hypothesized that the lunar craters formed either by (1) internal heat that melted and disrupted its surface crust (today we would describe this process as volcanism), or (2) impacts by particles from space (today these would be described as meteor impact craters). The controversy over these two origins for lunar craters actually continued until the 1970s, when it was finally resolved in favour of the impact hypothesis on the evidence of the lunar rocks returned by space missions.

Analogical reasoning was extensively employed by the geomorphologist Grove Karl Gilbert in his studies of (1) lunar craters, to which he correctly ascribed an impact origin (Gilbert, 1893), and (2) a crater in northern Arizona (now known as Meteor Crater; Plate 43), to which he incorrectly ascribed a volcanic origin (Gilbert 1896). The limitations of analogic reasoning in extraterrestrial geomorphology continue to the present day, as ably summarized by Mutch (1979):

1 Many landforms cannot be assigned a unique cause. Rather, the same landforms may be generated by different combinations of processes that converge to the same result.
2 The photointerpreter is artificially constrained in his analysis by his range of familiarity with natural landscapes. Because of these limitations, the student of extraterrestrial landforms must know as much as possible about the origin of landforms in general.

Planets, moons, and other objects

The term *planetary geomorphology* (Baker 1984) is also used for many of the topics covered by this

Plate 43 Meteor crater in northern Arizona. With a diameter of 1.2 kilometres, the crater formed about 25,000 years ago when an iron meteor struck the Earth at a velocity of about 11 kilometres per second

article, and it is true that planetary surfaces provide the sites of many landforms and landscapes (Greeley 1994). However, not all planets have rocky surfaces on which there are landforms. Moreover, there are many objects beyond Earth that are not planets, and many of these do indeed have landforms and landscapes. If we hold to the idea that one seeks to compare Earth's geomorphology to that of objects beyond Earth, then extraterrestrial geomorphology seems to be the appropriate term.

While future extraterrestrial geomorphology will surely extend to planetary objects in other solar systems, over a hundred of which have been discovered at the time of this writing, discussion here will be limited to the rocky objects of our own solar system. The inner planets, Mercury, Venus, Earth and Mars, all have rocky surfaces on which the effects of volcanism, tectonics and impact craters are in abundant evidence. Earth has a relatively large moon with a surface dominated by impact craters, the study of which has been directly facilitated by human visitation. Mars has two moons, but these are really captured asteroids, and are similar to many thousands of objects that occur throughout the inner solar system, mainly in the so-called 'asteroid belt' between Mars and Jupiter. The planets of the outer solar system, Jupiter, Saturn, Uranus and Neptune, are all giant gas balls, lacking any surface with landforms. Their satellites, however, are phenomenally rich in landscape complexity. Jupiter has four very large moons, Io, Europa, Ganymede, and Callisto, which form a kind of mini solar system, first discovered by Galileo's telescopic investigations. Ganymede and Callisto have heavily cratered surfaces on ice that is so cold it behaves as rock. Io's surface is dominated by volcanism that is much more active than any volcanism on Earth. Europa has a very young, nearly uncratered surface that is locally deformed because the icy crust of this moon overlies an immense ocean of liquid water. Other satellites of the outer planets are similar to asteroids in their surface character. Miranda, a moon of Uranus, looks to have been totally shattered by impact, and then accreted once more from the shattered remnants. The icy satellites of Uranus and Neptune are so cold that ices of ammonia and other volatiles comprise their rocky surfaces. A type of volcanism, known as 'cryovolcanism', was generated when these ices melted.

Titan, a moon of Saturn, has a diameter equal to about one-half that of Earth. It has an atmosphere that is slightly thicker than that of Earth, and, like Earth, the atmosphere is composed mainly of nitrogen. However, Titan is also extremely cold. The other main gas in its atmosphere is methane, and the great cold would lead to the condensation of that gas as a liquid on the surface of the satellite. Thus, Titan could have an ocean of methane and other hydrocarbons, or the liquid might just occupy lakes in the impact craters of a rocky surface, on which the 'rock' might be water ice. In any case, there is a very complex spacecraft, Cassini, which is scheduled to arrive in the Saturn system in July of 2004. The radar instrument on Cassini will permit penetration of the hazy atmosphere of Titan to reveal, for the first time to human observers, the landforms and landscapes of this haze-shrouded world.

Cratering and volcanism

Impact craters are the most ubiquitous landforms on the rocky and icy planets and satellites. Relatively low densities of impact craters indicate surfaces that are relatively young and unmodified relative to the 4.5 billion-year age of the solar system. Such surfaces comprise the dominant portions only of the large icy satellites Europa and Triton (a satellite of Neptune), the volcanically active satellite Io, and the planets Venus and Earth. In contrast, Mercury, Mars, the moon, and most planetary satellites have much of their surfaces covered with densely cratered terrains that formed over several billion years (Plate 44, p. 357). These surfaces are preserved because of minimal modification by active surficial processes related to atmospheric effects (exogenetic processes) and relatively localized volcanism and tectonic effects (endogenetic processes).

Volcanism is also very common in the solar system, though it does not dominate the landscapes of objects other than Io, Venus and Earth. All the rocky planets do have extensive volcanic plains, however. On Earth, these are hidden beneath ocean waters, and they were emplaced by seafloor spreading volcanism, mostly within the last 100 million years. Mercury has extensive intercrater plains, and both Mars and the moon have lowland plains that show evidence of being covered by lava flows. The lavas that formed these plains all seem to have been highly fluid, probably with basaltic compositions. Venus has some of the most extensive volcanic plains, and some of these are crossed by remarkable lava channels. The longest of these extends over 6,800 kilometres, making it longer than Earth's longest river (Baker *et al.* 1992).

Volcanic constructs, including large cones, shield volcanoes and calderas occur on Venus, Earth and Mars. Olympus Mons, a shield volcano on Mars, measures over 700 kilometres in diameter, and it rises to a height of 25 kilometres. It is only one example of extraterrestrial landforms that are much larger than their counterparts on Earth (Baker 1985). Though most extraterrestrial volcanic landforms are relict, the active volcanism of Io is spectacular. Eruptive plumes from the surface of Io were observed by the Voyager spacecraft to propel debris up to 300 kilometres above the surfaces and to deposit material up to 600 kilometres from the active vents. There are also active eruptive plumes on Triton, but the responsible process is probably more similar to that of a geyser than that of a volcano.

Tectonic landforms

Most of the rocky planet and satellite surfaces show evidence of structural deformation, with various fractures, graben and faults being the most common features. Mercury was deformed very early in its history by immense compressional forces that produced thrust fault landscapes. Mars has immense fracture zones and graben. However, only Earth exhibits the distinctive landforms associated with plate tectonics, including mid-oceanic ridges, transform faults and convergent continental margins with fold and thrust belt mountain ranges. Despite its density, radius and other geophysical similarities to Earth, the planet Venus does not show plate-tectonic landforms. This raises interesting questions about what makes plate tectonics unique to Earth.

Hillslopes and mass movement

Slopes occur on all the rocky planets. On airless bodies, only gravity and impact processes generate slope processes, but the atmospheres of Mars, Venus and Titan invite comparisons to other processes on Earth. A particularly interesting problem is the movement of extremely large (millions of cubic metres) slides or avalanches of rock and debris. Such masses on Earth have very high mobility over flat terrain. The cause of the very high mobility has been ascribed to the cushioning effect of air or water, reducing the effective pressure of the slide mass that would resist broad lateral spreading. However, these types of MASS MOVEMENT occur on the moon, which lacks both air and water. Many examples also occur on Mars, where air and water may have exerted influences.

Aeolian landscapes

While most extraterrestrial surfaces are airless, the atmospheres of Earth, Mars, Venus and Titan invite the interplanetary comparison of aeolian processes and landforms. Mars has the greatest variety of aeolian landforms. Crescent-shaped and transverse DUNES (see DUNE, AEOLIAN), wind ripples, YARDANGS, pitted and fluted rocks, and various dust streaks are all well displayed. There are also remarkable tracks produced by Martian dust devils. Aeolian bedforms also occur on Venus, which has an atmospheric pressure on the land surface that is ninety times that of Earth.

Channels, valleys and fluvial action

Besides Earth, fluvial action seems to have occurred only on Mars, and the most extensive fluvial processes were active in the remote past. The two main varieties of fluvial landforms on Mars are valley networks and outflow channels, morphological attributes of which are reviewed by Baker (1982, 2001). A great many of the valley networks occur in the old cratered highlands of Mars, leading to the view that nearly all of them formed during the heavy bombardment phase of planetary history, prior to 3 or 4 billion years ago. The outflow channels, in contrast, involve the immense upwelling of cataclysmic flood flows from subsurface sources, mostly during later episodes of Martian history. Much of the Martian surface is underlain by a thick ice-rich permafrost zone, a 'cryolithosphere', and the water feeding the outflow channels emerged from beneath this permafrost, possibly associated with volcanic processes (Baker 2001).

One of the most striking recent discoveries is that some water-related landforms on Mars are exceptionally young in age. This fact was prominently demonstrated by images from the Mars Global Surveyor orbiter showing numerous small gullies generated by surface runoff on hillslopes. The gullies were most likely formed by the melting of near-surface ground ice and the resulting debris-flow processes. The gullies are uncratered, and their associated debris-flow fan deposits are superimposed both on aeolian bedforms (dunes or wind ripples) and on polygonally patterned ground, all of which cover extensive areas that are also uncratered. Exceptionally young outflow channels and associated volcanism also occur on Mars. Data from Mars Global Surveyor show that localized water releases, interspersed with lava flows, occurred approximately within the last 10 million years. The huge discharges associated with these floods and the temporally related volcanism should have introduced considerable water into active hydrological circulation on Mars. It is tempting to hypothesize that the young outflow processes and volcanism are genetically related to other very young water-related landforms. The genetic connection for all these phenomena might well be climate change, induced by the water vapour and gases introduced to the atmosphere by both flooding and volcanism (Baker 2001)

Lakes, seas and 'oceans'

On Earth bodies of standing water include (1) lakes, in which the water is surrounded by extensive land areas, (2) seas, in which saline waters cover the greater part of the planetary surface, and (3) the ocean, which is the vast, interconnected body of water that covers about 70 per cent of Earth's surface. For Mars there is no direct geomorphological evidence that the majority of its surface was ever covered by standing water, though the term 'ocean' has been applied to temporary ancient inundations of the planet's northern plains. Although initially inferred from sedimentary landforms on the northern plains, inundation of the northern plains has been tied most controversially to identifications of 'shorelines'. New data indicate the presence of a regionally mantling layer of sediment, which seems to be contemporaneous with the huge ancient flood discharges of the outflow channels. Though the debate over the Martian 'ocean' has received much attention, even more compelling evidence supports the existence of numerous lakes, which were temporarily extant on the surface of Mars at various times in the planet's history.

Glacial and periglacial landforms

Evidence for past glacial activity on Mars is both abundant and controversial. The glacial features are also associated with periglacial landforms, which include debris flows, polygonally patterned ground, thermokarst, frost mounds, pingos and rock glaciers. On Earth most of these landforms develop under climate conditions that are both warmer and wetter than the conditions for cold-based glacial landforms (Baker 2001). The implications for past climatic change on Mars are profound because glaciers require substantial transport of atmospheric water vapour to sustain the snow accumulation that generates the positive mass balance needed for glacial growth.

The glacial landforms of Mars are erosional (grooves, streamlined/sculpted hills, drumlins, horns, cirques and tunnel valleys), depositional (eskers, moraines and kames), and ice-marginal (outwash plains, kettles and glacilacustrine plains). Of course, the landform names are all genetic designations, and ad hoc alternatives have been suggested for many. What is not ad hoc, however, is that all the glacial landforms occur in spatial associations, proximal-to-distal in regard to past ice margins, that would be obvious in a terrestrial setting. Areas of past glaciation on Mars (Kargel and Strom 1992) include the summits of very large volcanoes, uplands surrounding major impact basins (Plate 44), and the polar

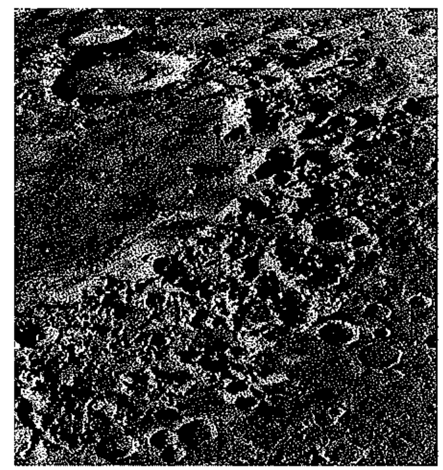


Plate 44 Oblique view of the Martian impact basin Argyre, surrounded by mountainous uplands (centre) many of which contain glacial features (Kargel and Strom 1992). Note the high clouds in the Martian atmosphere on the horizon

regions, where the ice caps were much more extensive during portions of post-Noachian time.

The future of geomorphology

It has long been apparent that the modern frontier of geomorphology, both as a matter of physical discovery and as an intellectual challenge, lies in the comparative study of planetary surfaces. This was summarized rather distinctly by Sharp (1980), as follows:

> Planetary exploration has proved to be a two-way street. It not only created interest in Earth-surface processes and features as analogues, it also caused terrestrial geologists to look at Earth for features and relationships better displayed on other planetary surfaces. Impact cratering, so extensive on Moon, Mercury, and Mars, is a well-known example. Another is the huge size of features, such as great landslides and widespread evidence of large-scale subsidence and collapse on Mars, which suggests that our thinking about features on Earth may have been too small-scaled. One of the lessons from space is to 'think big'.

References

Baker, V.R. (1982) *The Channels of Mars*, Austin, TX: University of Texas Press.

Baker, V.R. (1984) Planetary geomorphology, *Journal of Geological Education* 32, 236–246.

——(1985) Relief forms on planets, in A. Pitty (ed.) *Themes in Geomorphology*, 245–259, London: Croom Helm.

——(1993) Extraterrestrial geomorphology: science and philosophy of Earthlike planetary landscapes, *Geomorphology* 7, 9–35.

——(2001) Water and the Martian landscape, *Nature* 412, 228–236.

Baker, V.R., Komatsu, G., Parker, T.J., Kargel, J.S. and Lewis, J.S. (1992) Channels and valleys on Venus: preliminary analysis of Magellan data, *Journal of Geophysical Research* 97, 13,421–13,444.

Dorn, R.I. (2002) Analysis of geomorphology citations in the last quarter of the 20th century, *Earth Surface Processes and Landforms* 27, 667–672.

Gilbert, G.K. (1893) The Moon's face: a study of the origin of its features, *Philosophical Society of Washington Bulletin* 12, 241–292.

——(1896) The origins of hypotheses illustrated by the discussion of a topographic problem *Science*, n.s. 3, 1–13.

Greeley, R. (1994) *Planetary Landscapes*, Dordrecht, The Netherlands: Kluwer Academic.

Kargel, J.S. and Strom, R.G. (1992) Ancient glaciation on Mars, *Geology* 20, 3–7.

Mutch, T.A. (1979) Planetary surfaces, *Reviews of Geophysics and Space Physics* 17, 1,694–1,722.

Sharp, R.P. (1980) Geomorphological processes on terrestrial planetary surfaces, *Annual Review of Earth and Planetary Surfaces* 8, 231–261.

SEE ALSO: astrobleme; crater; geomorphology

VICTOR R. BAKER

F

FABRIC ANALYSIS

Measures one or more parameter of the three-dimensional disposition of elongated rock fragments in sediments. Such fragments have length, breadth and thickness, defined as a, b and c axes. The fragments contain three projection planes, maximum, intermediate and minimum. The maximum plane contains a and b axes, the intermediate a and c axes and the minimum b and c axes. Measurements of the orientation and dip values for axes and planes combined with statistical analysis can identify processes and environments of deposition for many sediment types (Andrews 1971; Dowdeswell and Sharp 1986). For example, in an undisturbed lodgement till the a-axes of pebbles are strongly oriented in the direction of local ice flow and dip slightly up-glacier. On the bed of a river cobbles and boulders frequently exhibit imbricate structure in which the a-axes are normal to the water current and the maximum plane dips upstream. In a storm beach deposit the a-axes of cobbles or shingles (flat cobbles) are usually deposited normal to the direction of wave advance and the maximum plane dips seaward.

References

Andrews, J.T. (1971) Methods in the analysis of till fabrics, in R.P. Goldthwait (ed.) *Till, A Symposium*, 321–327, Columbus: Ohio University Press.

Dowdeswell, J.A. and Sharp, M. (1986) Characterization of pebble fabrics in modern terrestrial glacigenic sediments, *Sedimentology* 33, 699–710.

ERIC A. COLHOUN

FACTOR OF SAFETY

The factor of safety, F, is defined as the ratio of the sum of resisting forces (shear strength) divided by the sum of driving forces (shear stress) of a slope:

$$F = \frac{\text{sum of resisting forces}}{\text{sum of driving forces}}$$

If at a location within a soil mass the shear stress becomes equal to the shear strength of the soil, failure will occur at that point. In this case $F = 1$. Where $F < 1$ the slope is in a condition for failure, where $F > 1$ the slope is likely to be stable.

Shear strength and shear stress were originally expressed by Coulomb in 1776. Shear strength of a soil is its maximum resistance to shear. Its value determines the stability of a slope. The knowledge of the shear strength is an essential prerequisite to any analysis of slope stability and the factor of safety. Coulomb postulated that:

$$\tau_f = c + \sigma \tan \varphi$$

where τ_f = maximum resistance to shear, c = cohesion of the soil, σ = total stress normal to the failure surface, and $\tan \varphi$ = angle of internal friction of the soil.

In 1925 Terzaghi published the fundamental concept of effective stress, $\sigma' = \sigma - u$ (with u = pore-water pressure), that water cannot sustain shear stress and that shear stress in a soil can be resisted only by the skeleton of solid particles at the particle contact points. Shear strength is expressed as a function of effective normal stress as:

$$\tau_f = c' + \sigma' \tan \varphi'$$

in which the parameters c′ (effective cohesion) and φ' (effective angle of friction) are properties of the soil skeleton.

The factor of safety can then be expressed as

$$F = \frac{c' + \sigma' \tan \varphi'}{\tau_f}$$

This equation can be used for limit equilibrium methods in slope stability analysis (Duncan 1996). The calculation of F requires the description of a potential slip surface which is defined as a mechanical idealization of the failure surface. The critical slip surface is the one with the minimum value of F of all possible slip surfaces included in the limit equilibrium calculation.

Most natural hillslopes prone to landsliding have F values between about 1 and 1.3, 'but such estimates depend upon an accurate knowledge of all the forces involved and for practical purposes design engineers always adopt very conservative estimates of stability' (Selby 1993). In practice the highest uncertainties are related to soil water, especially with the spatial variability of PORE-WATER PRESSURE and seepage.

In geomorphology the factor of safety concept is essential to understand landscape stability. F is considered as ratio of landform strength resistence and the magnitude of impacting forces.

References

Duncan, J.M. (1996) Soil Slope Analysis, in A.K. Turner and R.L. Schuster (ed.) *Landslides. Investigation and Mitigation*, Transportation Research Board, Special Report 247, 36–75, Washington, DC: National Academy Press.

Selby, M.J. (1993) *Hillslope Materials and Processes*, Oxford: Oxford University Press.

Further reading

Craig, R.F. (1994) *Soil Mechanics*, London: Chapman and Hall.

SEE ALSO: shear and shear surface

RICHARD DIKAU

FAILURE

Within geotechnical geomorphology, the term failure implies the occurrence of a dislocation within a material, usually accompanied by detachment of a soil or rock mass. The most common case involves shear (see SHEAR AND SHEAR SURFACE) failure along a well-defined plane of rupture as a LANDSLIDE. In hard sedimentary strata, igneous and metamorphic rocks, detachments usually occur along planes of weakness defined by bedding, joints (see JOINTING), foliation and faults. The potential for movement is greatest where layers dip downslope. The resultant translational landslides are called rockslides. For dry layers, the kinematic criterion for failure is:

$$\phi' < \delta < \beta$$

where ϕ' is the friction angle along the layers, δ is the dip angle, and β the slope angle. The inequality shows that the weak layer must crop out on the slope (i.e. $\delta < \beta$), and the dip angle must exceed the friction angle. In most cases, the temporal variation of all three parameters is typically small. However, frictional resistance along joints can be abruptly reduced when water pressure increases. The steepest stable angle, θ_c, of a rock layer is then:

$$\theta_c = [(\gamma_{sat} - m\gamma_w) / \gamma_{sat}] \tan \phi'$$

where γ_{sat} is the saturated unit weight of the material, γ_w is the unit weight of water, and parameter m is the ratio of the saturated depth to the total slab depth. Similar principles apply to failures in colluvial materials where planar detachments called debris slides often occur at the interface between COLLUVIUM and harder materials below, such as bedrock. Debris slides also occur in glacial materials, for example at the interface between loose unconsolidated ablation till and denser basal till below. In softer rocks, such as clays, shales and mudstones, a lesser degree of structural control exists, and shear surfaces often run oblique to the direction of bedding as rotational failures. In these cases, the above assumption of a plane translational slide is no longer valid.

A more general way to assess the stability or proximity to failure of a soil or rock mass, of any geometrical shape, is the Mohr–Coulomb equation:

$$s = c' + (\sigma - u) \tan \phi'$$

where s is shear strength, c′ is COHESION, σ is total normal stress, u is PORE-WATER PRESSURE and ϕ' the angle of shearing resistance. Parameters c′ and ϕ' are material properties that control shear strength at the ambient EFFECTIVE STRESS. Pore pressure, u, is independent of these parameters, and is a function of moisture recharge from antecedent and ambient climatic events. The overall stability of a mass is assessed from its FACTOR OF SAFETY, $F = s/\tau$, where τ is shear stress. By definition, failure occurs when F = 1.0. Failure is most commonly caused by saturation, which causes an increase in shear stress concomitant with a reduction in frictional strength. This explains why so many landslides are associated with major rainstorms or snowmelt.

The above version of the Mohr–Coulomb equation applies to drained failures, which involve no excess pore pressures. In the case of rapid movements in low density, fine-grained, saturated soils (for example, QUICKCLAYS), collapse of the soil structure under shear loads causes significant excess pore pressures to develop. Such failures involve LIQUEFACTION, and must be analysed with reference to undrained (see UNDRAINED LOADING) strength parameters.

Failure may also occur by toppling and buckling of layers, especially in thinly bedded rocks. Toppling involves forward and downslope rotation of layers and is common where strata dip steeply into a slope. For single blocks the toppling criterion is:

$$b/h < \tan \delta$$

where b and h are the breadth and height of the block and δ is the inclination of the block's base. For flexural toppling, which involves downslope rotation and interlayer slip, the criterion is:

$$\alpha < \beta - \phi'$$

where pole angle $\alpha = (90° - \delta)$ is the angle of the normal to the plane, and δ is the dip angle. The inequality shows that toppling is most likely to occur in steeply dipping strata, but may be enhanced where slopes are undercut and steepened. Buckling tends to occur where ductile, thinly bedded rocks, such as argillite and phyllite, dip downslope slightly steeper than the slope angle. When the downslope compressive stress exceeds the bending resistance of the layers, buckling may occur.

Most slope failures involve more than one type of movement. For example, a landslide dominated by plane failure at its base may also involve buckling or forward toppling of material by compression at the toe area, and tensional failure at the headscarp. Transitions from one type of movement to another are also common, for example detachment of a saturated mass as a debris slide, followed by disintegration and fluidization as a DEBRIS FLOW further downslope.

Although the Mohr–Coulomb equation implies abrupt attainment of failure, many landslides probably involve slow creep movements prior to detachment. Deep-seated gravitational movements of the SACKUNG type probably involve prolonged, slow movements at depth. Such mountain scale masses total tens to hundreds of millions of cubic metres of material moving at millimetres to centimetres per year. The surface expression of such movements is typically tension cracks, uphill facing scarps and grabens, or is less clearly defined as masses of broken, dilated rock. Such movements may occur over centuries to millennia without the development of a landslide rupture surface. However, other cases are known to have terminated in large rock avalanches (STURZSTROMS). This suggests that a continuum of slope movement rates and types may occur over time at an individual site, a circumstance which is not easily encompassed by existing methods used to classify and analyse slope movements.

Further reading

Barnes, G.E. (2000) *Soil Mechanics*, 2nd edition, London: Macmillan.

Bovis, M.J. and Evans, S.G. (1996) Extensive deformations of rock slopes in the southern Coast Mountains, British Columbia, Canada, *Engineering Geology* 44, 163–182.

Goodman, R.E. and Bray, J.W. (1976) Toppling of rock slopes, in *Proceedings of a Specialty Conference on Rock Engineering for Foundations and Slopes*, Boulder, CO; New York: American Association of Civil Engineers.

Hoek, E. and Bray, J.W. (1981) *Rock Slope Engineering*, 3rd edition, London: Institution of Mining and Metallurgy.

Turner, A.K. and Schuster, R.L. (eds) (1996) *Landslides: Investigation and Mitigation*, Washington, DC: National Academy Press.

MICHAEL J. BOVIS

FALL LINE

The topographical and geological boundary between an upland region of relatively high resistance crystalline rock and a lower region of weaker rock. Rivers transcending this boundary often develop waterfalls and rapids in parallel. A less frequent and appropriate use of the term is for the point where a river ceases to be tidal. The fall line is thus a geological and geomorphological boundary. Generally, streams and rivers upstream from the fall line have small floodplains and are of low sinuosity, whereas downstream from the fall line rivers and streams tend to possess larger floodplains and display high sinuosity.

The type example of a fall line is the eastern United States region, where the upland Piedmont Plateau (crystalline rock) meets the Atlantic coastal plain (weaker sedimentary rock). The junction is marked by rapids and waterfalls on each of the major rivers that transcend the zone (i.e. the Delaware, Potomac, James, Savannah, etc.).

The steep gradient of the American fall line has been accounted for in three main ways, as reviewed by Renner (1927) and Lobeck (1930: 454). First, the feature can be interpreted as a zone of monoclinal flexing or faulting (though faulting occurs on the fall line in few localities). Second, as an area where the rivers have eroded away the softer rocks of the coastal plain, forming knickpoints at the boundary with the resistant crystalline piedmont rocks. Third, as the intersection of two ancient peneplains, in which the older mid-Mesozoic erosion surface plunges beneath the coastal plain deposits that overlie the younger peneplain. The fall line represents a stripped part of the older peneplain and accounts for its steeper slope.

References

Lobeck, A.K. (1930) *Geomorphology: An Introduction to the Study of Landscapes*, London and New York: McGraw-Hill.

Renner, G.T. Jr (1927) The physiographic interpretation of the Fall Line, *Geographical Review* 17(2), 278–286.

STEVE WARD

FANGLOMERATE

A sedimentary rock consisting of heterogeneous fragments of assorted size deposited in an alluvial fan and subsequently cemented into a solid mass. The term was introduced by Lawson (1913) to describe the coarse upslope parts of ALLUVIAL FAN formations, though the term is also used more generally for conglomerates and breccias deposited on alluvial fans. They are composed of two main facies: water laid deposits and mass flow deposits. Fanglomerates are characterized by their parallel bedding and decreasing particle size downslope, alongside rapid fan thinning.

References

Lawson, A.C. (1913) *The Petrographic Designation of Alluvial Fan Formations*, University of California Publications Department G.7, 325–334.

STEVE WARD

FAULT AND FAULT SCARP

A *fault* is a surface or zone along which one side has moved relative to the other in a direction parallel to the surface or zone. The term is applied to features extending over distances of metres or larger, whereas those at the scale of centimetres are called shear fractures, and those at the scale of millimetres are microfaults. Most of them are brittle structures, although some may represent ductile deformation. For an inclined fault, the fault block above the fault is the hanging wall, and the block below the fault is the footwall. Faults are subdivided into dip-slip (showing slip parallel to the dip of the fault surface), strike-slip (of slip parallel to the strike of the fault surface), and oblique-slip faults (where slip is inclined obliquely on the fault surface). The dip-slip faults include normal faults, on which the hanging wall block moves down relative to the footwall block, and thrust (dipping $<45°$) or reverse faults (dipping $>45°$), on which the hanging wall moves up relative to the footwall block. Strike-slip faults are either right-lateral or left-lateral, depending on the sense of motion of the fault block across the fault from the observer. These faults are commonly planar and vertical. Both dip-slip and strike-slip faults frequently form a linked fault system which, in cross section, consists of flats and ramps, which cut through the footwall and detach a slice of hanging wall rocks.

Some normal faults are concave-upward faults of dip decreasing with increasing depth; they are called listric faults. These can join or turn into a low-angle detachment fault at depth. Small-scale faults parallel to the major fault and showing the same sense of shear are called synthetic faults; those of the conjugate orientation are antithetic faults. A downthrown block bounded on either side by conjugate normal faults is a graben, whereas a relatively uplifted block bounded by two conjugate faults is a horst. A half-graben is a lowered tilted block bounded on one side by a normal fault. Step faults are parallel faults on which the downthrown side is on the same side of

each fault. Rotational movements between the two fault blocks result in varying throws along the fault strike, producing either hinge faults, where displacement increases from zero to a maximum along the strike, or pivot (scissor) faults, where one block appears to have rotated about a point on the fault plane. The traces of normal faults are either straight or slightly sinuous, depending on the fault dip, whereas the traces of thrusts are usually highly sinuous due to low-angle intersection with the ground surface.

Large-scale strike-slip faults are called transform faults when building segments of lithospheric plate boundaries, or trancurrent faults when they occur in continental crust and are not parts of plate margins. Strike-slip faults frequently form bends (curved parts of the fault trace) and stepovers, i.e. places where one fault ends and another, *en echelon* fault begins. A left bend or stepover in a right-lateral fault system (restraining bend) induces local compression (uplift; transpression), whereas a right bend or stepover in a right-lateral fault (releasing bend) produces local extension (subsidence; transtension). Displacement at extensional bends and stepovers forms rhomboidal, fault-bounded depressions, called pull-apart basins.

A *fault scarp* is a tectonic landform coincident with a fault plane that has displaced the ground surface. A residual fault scarp is a mature scarp, upon which the original tectonic surface has been obliterated by geomorphic processes. A fault-line scarp, in turn, results from differential weathering and erosion of the rocks on either side of the fault.

Scarps produced by normal faulting are usually located at the contact between bedrock in the footwall and Quaternary sediments in the hanging wall. Scarps associated with reverse faulting in solid bedrock are commonly overhanging and tend to collapse and/or be eroded; they are also more deeply embayed than their normal counterparts. Scarps associated with strike-slip faults are less prominent and are best developed in areas of uneven topography. In loose sediments, however, fold-limb (monoclinal or fold) scarps are formed, and usually are surface expression of blind thrusts.

Active normal fault scarps include: piedmont (simple) scarps, formed in unconsolidated deposits; multiple (complex) scarps, related to formation of a fault splay during a single faulting event; composite (multi-event, compound) scarps, up to a few tens of metres high, formed due to renewed slip on a fault; and splintered scarps, produced due to fault displacement distributed across overlapping *en echelon* segments. A piedmont scarp (Wallace 1977) includes a steep ($>50°$) free face, a moderately inclined (30–40°) debris slope, and a gently inclined (5–10°) wash slope (see SEISMOTECTONIC GEOMORPHOLOGY). Fault scarps in semi-arid climate degrade from gravity-controlled (10^2 yrs), through debris-controlled (10^3 yrs), to wash-controlled (10^5 yrs) slope due to either: decline, replacement, retreat or rounding (Mayer 1986).

Depending on the climatically controlled rate of removal of debris shed from the scarp, the Oregon or Basin and Range-type scarps, typical of semi-arid climate, and the Awatere or New Zealand-type scarps, formed in a more humid climate, have been distinguished. Faulting in bedrock is accompanied by fracturing and brecciation which can seriously modify the bedrock susceptibility to erosion. Fault rocks of contrasting resistance to erosion are typical of the Aegean-type fault scarps (Stewart and Hancock 1988), where normal faults in carbonate bedrock are underlain by different types of alternating compact and incohesive breccias. Degradation of such scarps proceeds differently as compared to the Nevada-type model of piedmont scarp.

Colluvial wedges shed from fault scarps can be dated by: ^{14}C, luminescence, dendrochronological, palynological, tephrochronological and weathering rates techniques. Scarps formed in loose sediments can be modelled mathematically by: linear regression, diffusion modelling and statistical analysis of scarp parameters.

Due to repeated episodes of faulting, bedrock fault escarpments, several hundred metres high, and fault-generated range fronts, several hundreds of kilometres long and up to 1 km high, can form. The range front morphology is determined mainly by the ratio of uplift to erosion. Range fronts in a humid climate may appear more degraded than range fronts with the same uplift rate in an arid climate. Active normal fault-generated mountain fronts frequently display triangular or trapezoidal facets (faceted spurs, flat irons) that form due to uplift and dissection of a normal scarp by gullies and whose bases are parallel to the fault trace. Flights of faceted spurs have been interpreted as a result of either episodic uplift, distributed faulting within the range-bounding fault, or even active landsliding.

References

Mayer, L. (1986) Tectonic geomorphology of escarpments and mountain fronts, in R.E. Wallace (ed.) *Active Tectonics*, 125–35, Washington, DC: National Academy Press.

Stewart, I.S. and Hancock, P.L. (1988) Fault zone evolution and fault scarp degradation in the Aegean region, *Basin Research* 1, 139–152.

Wallace, R.E. (1977) Profiles and ages of young fault scarps, north-central Nevada, *Geological Society of America Bulletin* 88, 1,267–1,281.

Further reading

Bloom, A.L. (1978) *Geomorphology*, Englewood Cliffs, NJ: Prentice Hall.

Burbank, D.W. and Anderson, R.S. (2001) *Tectonic Geomorphology*, Malden: Blackwell.

Cotton, C.A. (1958) *Geomorphology*, Christchurch: Whitcombe and Tombs.

Keller, E.A. and Pinter, N. (1996) *Active Tectonics*, Upper Saddle River, NJ: Prentice Hall.

McCalpin, J.P. (ed.) (1996) *Paleoseismology*, San Diego: Academic Press.

Morisawa, M. and Hack, J.T. (eds) (1985) *Tectonic Geomorphology*, Boston: Allen and Unwin.

Stewart, I.S. and Hancock, P.L. (1990) What is a fault scarp?, *Episodes* 13, 256–263.

Stewart, I.S. and Hancock, P.L. (1994) Neotectonics, in P.L. Hancock (ed.) *Continental Deformation*, 370–409, London: Pergamon Press.

Turner, J.P. (2000) Faults and faulting, in P.L. Hancock and B.J. Skinner (eds) *The Oxford Companion to the Earth*, 342–345, Oxford: Oxford University Press.

Twiss, R.J. and Moores, E.M. (1992) *Structural Geology*, New York: W.H. Freeman.

SEE ALSO: seismotectonic geomorphology

WITOLD ZUCHIEWICZ

FECH-FECH

A term applied in the Sahara Desert to fine silt with a powdery consistency and to fine superficial deposits with a low density that often contain evaporates. Progress across fech-fech is made difficult by the absence of cohesion of particles (where feet or wheels penetrate). Fech-fech can be classified from a genetic point of view into two main types:

- Fech-fech developed on Holocene lacustrine muds or fluvio-lacustrine sediments: soft zones within the lacustrine limestones, with 40 per cent of fine particles ($<20\mu$) and a higher content of soluble salts, differentiate these sediments in an environment which is almost exclusively sandy.
- Fech-fech developed on clayey shales: the present-day weathering of shales leads to their superficial expansion, which is accentuated by the incorporation of aeolian detrital particles between the disconnected layers.

In addition to these two types, one also finds fech-fech on Quaternary regs with a denser structure ($1.5\,g\,cm^{-3}$) due to the formation of aggregates of silty and salty sand. Tracks can be preserved for a long time on these soft regs where they are underlain by a sandy layer.

Further reading

Conrad, G. (1969) *L'évolution continentale post-hercynienne du Sahara algérien (Saoura, Erg Chech, Tanezrouft, Ahnet-Mouydir)*, série: Géologie No. 10, Paris: Centre National de la Recherche Scientifique (CNRS).

MOHAMED TAHAR BENAZZOUZ

FERRALLITIZATION

A process characterized by the aggressive leaching of a substrate as a consequence of intense tropical weathering, and in which the net effect is a relative accumulation of iron-rich (and commonly also aluminium-rich) compounds; in particular iron and alumina sesquioxides. Ferrallitization is an *in situ* process during which prolonged or intense weathering causes the breakdown of the primary constituents of a pre-existing soil or rock substrate.

Ferrallitization progresses upon rock substrates by the hydrolysis of their primary minerals. This leads to the individualization of the chemical elements of these minerals, the complete leaching of constituent alkali and alkali Earth elements, and the partial or total leaching of silica. Once breakdown commences and elements released, they become available for removal from the system, whilst less mobile constituents, such as Fe, Al, and Ti, remain behind as residual materials and form sesquioxides. Since these less mobile constituents are present in significant proportions within many common substrate lithologies (e.g. 12–18 per cent Fe_2O_3, 12–18 per cent Al_2O_3 in continental basalt, and 0.5–5 per cent Fe_2O_3, 12–15 per cent Al_2O_3 in granitic materials), the residuum readily becomes enriched in Fe and Al. Any silica remaining in the residuum is present either as corroded primary quartz crystals or grains, or else becomes combined into alteration products (e.g. kaolinite and gibbsite). Since, by definition, laterites (see FERRICRETE) form by *in situ* mineral breakdown, ferrallitization represents a key process in the development of lateritic weathering profiles.

MIKE WIDDOWSON

FERRICRETE

A horizon, at the land surface, made up of the cementation of near-surface materials by iron oxides, and often forming a resistant DURICRUST. Typically between 1–20 m in thickness, it can form laterally extensive sheets which may extend over a few, to hundreds, or even thousands of km². Consequently, it is perhaps the most widespread of all the duricrust materials. At outcrop it comprises a massive, interlocking fretwork of iron, and often aluminium compounds (i.e. sesquioxides) that bind together other lithological and pedogenic components.

Ferricrete has a long history of study by geologists, geomorphologists, pedologists and agronomists. Considerable effort has been directed toward determining the conditions under which it forms, and this has proved crucial in advancing many aspects of TROPICAL GEOMORPHOLOGY (Thomas 1994; Widdowson 1997). Moreover, chemical and physical durability of ferricrete has meant that it has often played a prominent role in evolution of tropical, and subtropical landscapes (e.g. McFarlane 1971; Bowden 1987; Widdowson and Cox 1996).

In its broadest sense, the term ferricrete can be used to describe any duricrust material in which the dominant bulk components are iron-rich compounds. However, whilst this may seem a straightforward definition, difficulties arise because the term has been employed to describe a wide range of terrigenous weathering and alteration products resulting from differing processes of formation (Ollier and Galloway 1990). Therefore, it becomes important to understand the differences and, where possible, make distinctions between genetically different types of iron-rich duricrust.

Since the nineteenth and early twentieth centuries, the terms ferricrete ('an iron-rich crust'; Lamplugh 1907) and laterite ('a highly weathered material rich in secondary forms of iron and/or aluminium'; (Buchanan 1807; Babbington 1821; Sivarajasingham *et al.* 1962; Plate 45) have been used interchangeably to describe iron-rich duricrusts of various genetic origins. This has led to considerable confusion. However, the problems of co-ordinating laterite and ferricrete description stem not only from investigation by a variety of different scientific disciplines, but also from the development of extensive anglophone and francophone descriptive terminologies. Nevertheless, it is evident from field studies that the majority of iron-rich duricrusts can be adequately described in terms of two genetically distinct types. Aleva (1994) distinguishes between those duricrusts in which an absolute iron enrichment occurs (i.e. those which receive a net input of iron), and those which attain their elevated iron contents through residual enrichment within the profile (i.e. no net input of iron).

Ferricretes are those duricrusts which incorporate materials non-indigenous to the immediate locality in which the duricrust formed. In many instances the transported materials can be readily

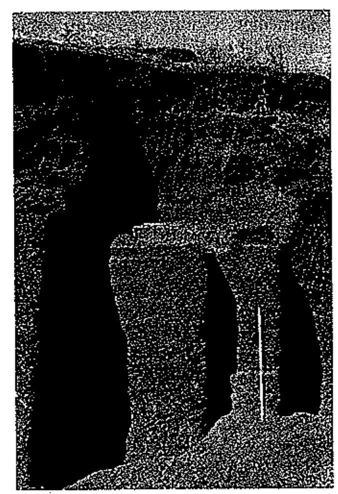


Plate 45 Laterite quarry near Bidar, south-east Deccan, India (with 1 m scale in lower right). Material beneath the indurated duricrust of a laterite profile is excavated, cut into large bricks, and allowed to harden in the sun. This is similar to the material first named 'laterite' by Buchanan (1807)

identified as pebbles or clasts derived from adjacent lithological terranes, or as fragments from indurated layers of earlier generations of laterite or ferricrete (Plate 46). Importantly, the term ferricrete should also be extended to those materials whose constituents have been substantially augmented by the precipitation or capture of elements and compounds from allochthonous fluids (i.e. those derived during the breakdown and mobilization of materials outside the immediate locality of ferricrete formation). Although it is the allochthony of the constituent materials of the ferricrete which justify its appellation, determining whether the introduction of such fluids has taken place, and confirming their allochthony, is often problematic. However, since ferricretes may develop as ferruginous foot slope accumulations or within topographic depressions, they can often be distinguished by the fact that they display an obvious discordance with the underlying substrate lithologies. In effect, they do not display the progressive weathering profile characteristic of many laterite profiles, and instead the ferricrete horizon sits upon relatively unaltered bedrock.

Laterites are iron-rich duricrusts which have formed directly from the breakdown of materials in their immediate vicinity, and so do not contain any readily identifiable allochthonous component. Lateritic duricrusts are typically manifest as the uppermost layers of *in situ* weathering profiles. Where these profiles are fully exposed, such as the widespread examples developed on basalt in western India (Widdowson and Gunnell 1999), they consist of an uninterrupted progression from unaltered bedrock, through the WEATHERING FRONT into SAPROLITE (in which structure and crystal pseudomorphs of the parent rock may still be recognized), and then upward through increasingly altered and iron-enriched zones that culminate as a highly indurated 'tubular' laterite at the top of the profile (Plate 47).

To summarize, ferricrete and laterite are not synonymous terms and should, wherever possible, be used to distinguish between fundamentally different types of iron-rich duricrust. This distinction is particularly important since it places constraints upon the type of processes operating during evolution of a duricrust, and the palaeoclimatic and morphological conditions existing at the time of its development. However, although emphasis is put upon establishing whether the iron component is allochthonous or autochthonous, distinguishing these two types of duricrust, both in the field and in hand specimen, can prove problematic. Problems arise because, once formed, ferricretes can begin to alter and evolve in response to prevailing climatic and groundwater conditions (Bowden 1997) and, over time, begin to exhibit some of the structural and textural features typical of lateritic weathering profiles. In effect these

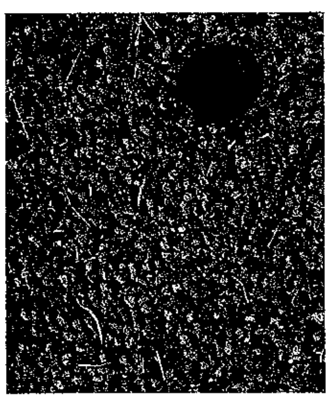


Plate 46 Granular ferricrete surface comprising allochthonous materials derived from earlier generations of laterite and ferricrete, near Bunbury, Western Australia

Plate 47 Indurated 'tubular' laterite sample from the top of an *in situ* weathering profile near Bunbury, Western Australia

'evolved' ferricretes become modified by a post-depositional weathering and ferrallitization overprint. Conversely, the role of allochthonous groundwater fluids, and associated lateral or downslope transport of elements and compounds, cannot always be excluded in the development of otherwise autochthonous laterite weathering profiles.

More recently, ferricrete and laterite duricrusts, together with Fe-rich palaeosols, have begun to acquire renewed importance as palaeoenvironmental indicators (e.g. Bardossy 1981; Thomas 1994; Tsekhovskii *et al.* 1995). The investigation of such materials within the geological record, together with appropriate mineralogical, geochemical and isotopic studies, can now reveal detailed information regarding past climatic and atmospheric conditions. For instance, geochemical and isotopic analyses of Proterozoic laterites from South Africa (Gutzmer and Beukes 1998), suggest not only an ancient oxidizing atmosphere, but also a hot and humid climate at *c.*2 Ga. Moreover, carbon isotope signatures preserved within these laterites may indicate the presence of an early terrestrial vegetation.

References

Aleva, G.J.J. (compiler) (1994) *Laterites. Concepts, Geology, Morphology and Chemistry*, Wageningen: ISRIC.

Babbington, B. (1821) Remarks on the geology of the country between Tellicherry and Madras, *Transactions of the Geological Society of London* 5, 328–329.

Bardossy, G. (1981) Palaeoenvironments of laterites and lateritic bauxites – effect of global tectonism on bauxite formation, *Proceedings of the International Seminar on Lateritisation Processes* (Trivandrum, India, 11–14 December 1979), 287–294, Rotterdam: Balkema.

Bowden, D.J. (1987) On the composition and fabric of the footslope laterites (duricrust) of Sierra Leone, West Africa, and their geomorpholoical significance, *Zeitschrift für Geomorphologie NF, Supplementband* 64, 39–53.

—— (1997) The geochemistry and development of lateritized footslope benches: the Kasewe Hills, Sierra Leone, in M. Widdowson (ed.) *Palaeosurfaces: Recognition, Reconstruction, and Paleoenvironmental Interpretation*, Geological Society of London Special Publication 120, 295–306.

Buchanan, F. (1807) A journey from Madras through the countries of Mysore, Kanara, and Malabar Vol. 2, 436–461; Vol. 3, 66, 89, 251, 258, 378, London: East India Co.

Gutzmer, J. and Beukes, N.J. (1998) Earliest laterites and possible evidence for terrestrial vegetation in the Early Proterozoic, *Geology* 26, 263–266.

Lamplugh, G.W. (1907) Geology of the Zambezi basin around Batoka Gorge, *Quarterly Journal of the Geological Society of London* 63, 162–216.

McFarlane, M.J. (1971) Lateritization and landscape development in Kyagwe, Uganda, *Quarterly Journal of the Geological Society of London* 126, 501–539.

Ollier, C.D. and Galloway, R.W. (1990) The laterite profile, ferricrete and unconformity, *Catena* 17, 97–109.

Sivarajasingham, S., Alexander, L.T., Cady, J.G. and Cline, M.G. (1962) Laterite, *Advances in Agronomy* 14, 1–60.

Thomas, M.F. (1994) *Geomorphology in the Tropics*, Chichester: Wiley.

Tsekhovskii, Yu G., Shchipakina, I.G. and Khramtsov, I.N. (1995) Lateritic eluvium and its redeposition products as indicators of Aptian–Turonian climate. *Stratigraphy and Geological Correlation* 3(3), 285–294.

Widdowson, M. (ed.) (1997) *Palaeosurfaces: Recognition, Reconstruction, and Paleoenvironmental Interpretation*, Geological Society of London Special Publication 120.

Widdowson, M. and Cox, K.G. (1996) Uplift and erosional history of the Deccan traps, India: evidence from laterites and drainage patterns of the Western Ghats and Konkan Coast, *Earth and Planetary Science Letters* 137, 57–69.

Widdowson, M. and Gunnell, Y. (1999) Lateritization, geomorphology and geodynamics of a passive continental margin: the Konkan and Kanara lowlands of western peninsular India. Special Publication of the the International Association of Sedimentologists 27, 245–274.

SEE ALSO: duricrust

MIKE WIDDOWSON

FIRE

Wildfire is one of the most potent agents of geomorphic change, modifying processes and greatly increasing erosion and deposition rates in virtually all BIOGEOMORPHOLOGY environments. This entry covers the geomorphological role of fire in WEATHERING, soils, HILLSLOPE PROCESSES and river systems.

Although most texts and reviews list fire as an important weathering agent (Blackwelder 1927), most field studies are anecdotal or supported by little data with fewer experimental controls. Thus, the most important insights on fire weathering result from laboratory studies (Goudie *et al.* 1992) where experimentalists have learned that fire weathering depends heavily on rock physical properties, varies with different rock types, is faster in smaller rocks, and fire weathering rates increase with increasing water content.

An example of the impact of fire on rock weathering comes from the April–May 2000 'Coon Creek' wildfire that burned around 37.5 km^2 of the Sierra Ancha Mountains, 32.3 km north of Globe, Arizona – including 25 sandstone and 19 diorite boulders surveyed in 1989 and resurveyed (a) after the burn, (b) after the summer 2000 precipitation season, and again (c) after the winter 2001 snow season (Dorn 2003, Plate 48). When stretched over cumulative boulder areas, erosion immediately after this single fire averaged >26 mm for sandstone and > 42 mm for diorite. But averages are misleading, because sandstone and diorite boulders expressed bimodal patterns of erosion, where fire-induced weathering generated either (a) no

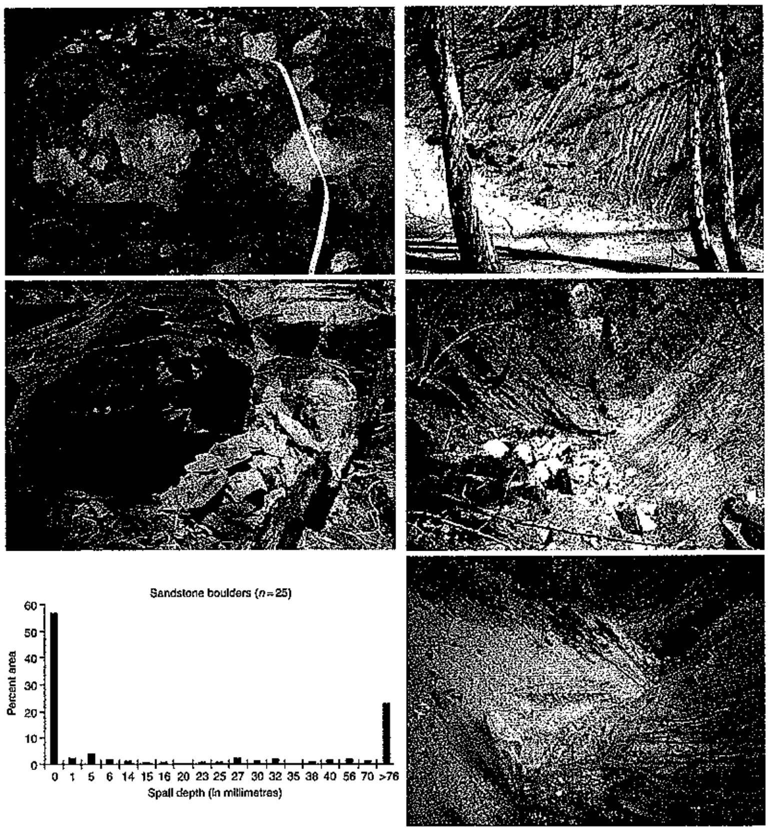


Plate 48 Left column: boulder weathering from the Coon Creek Spring 2000 fire, Sierra Ancha Mountains, Arizona. The top left image shows flaking of millimetre-scale spalls. The centre-left image shows where half of a boulder fragmented as a result of the fire. The graph on the lower left shows the overall bimodal pattern, whereby fire weathering produces erosion of small flakes or extensive slabs. Right column: fire-generated erosion from the 1995 Storm King fire, Colorado, courtesy of the US Geological Survey (Canon *et al.* 2001; see also http://landslides.usgs.gov/html_files/ofr95-508/index.html). The upper right image shows post-fire rill erosion. Other photos show in-channel conditions before (middle right) and after (lower right) passage of a debris flow

erosion to thin, millimetre-scale spalling or (b) massive spalls thicker than 7.6 cm. This field study confirms an earlier experimental finding that fire increases a rock's susceptibility to post-fire weathering and erosion processes (Goudie *et al.* 1992), since summer-time convective storms and subsequent winter snows continued to promote boulder erosion on the order of 1–5 millimetres. In addition to erosion of boulder surfaces, 85-metre- diameter boulders appear to have been fragmented into cm-scale clasts – suggesting that fire can modify hillslope evolution in locations where boulders are important controls on the evolution of slopes.

Wildfire generates extensive changes to soil systems (Morris and Moses 1987), perhaps the most important being the development of HYDROPHOBIC SOILS. Wildfires produce volatile hydrocarbons that penetrate soil up to 15 centimetres and make a water-repellent layer. In addition, fire ash decreases the ability of soils to adsorb water. Field checks involve digging a progressively deeper trench and applying water. Water that does not infiltrate immediately (within 10 seconds) indicates the soil is hydrophobic. Extreme hydrophobicity results in water ponding for more than 30 seconds. On unburned slopes, normal biogeomorphology processes decrease soil erosion, for example, by intercepting raindrop impacts, increasing infiltration and providing structural support. Hydrophobicity from burning decreases infiltration capacity, and increases OVERLAND FLOW and SOIL EROSION.

Even before it rains, burning enhances erosion by dust DEFLATION and dry ravel. Dry ravel is a type of granular MASS MOVEMENT where frictional and collisional particle interactions dominate flow behaviour, all not requiring rainfall. Dry ravel provides sediment to channels from particularly steep slopes, and this process is well documented after southern California fires.

Burning greatly increases surface runoff from precipitation, which increases the volume and velocity of the surface runoff. Higher discharge of surface water flows then result in the formation of RILLS and gullies (see GULLY) on hillsides. Fire-enhanced gullies and rills transport surface runoff and sediment to stream channels. Peak flows in the channel tend to occur with less of a lag time than those observed in unburned watersheds. Flood peaks tend to be much higher and more capable of eroding sediment stored in channels, leading to channel incision.

The sediment load of the fluvial system also changes after a fire. Sediment from a number of different sources may be incorporated into flows progressing down a hillside or channel. Sediment-water flows on burned slopes change the concentration, size distribution and/or composition of the entrained sediment to the point where a change in measurable yield strength takes place; this change is called HYPERCONCENTRATED FLOW. In hyperconcentrated flows, particles are deposited as individual grains from suspension, and the remaining fluid continues to move.

Fires also greatly increase DEBRIS FLOWS (Cannon *et al.* 2001; Swanson 1981). In contrast to streamflow or hyperconcentrated flow, debris flows host a sediment-water mixture that moves as a single phase. Deposition does not separate out particles, so debris-flow deposits have sharp, well-defined flow boundaries. The most recognizable deposits are levees lining flow paths and lobes of material at a flow terminus. Many terms have been used for the processes and deposits of debris flows, including slurry flow, mudflow and debris torrent.

Fire-enhanced debris flows start by landsliding or sediment bulking of surface water flows. Landsliding after burning tends to be more common in colluvial-filled hollows on slopes, where unconsolidated thick deposits of colluvium fail after rainfall. This landslide then mobilizes into a debris flow, where the debris-flow path can then be traced up to a landslide-scar source.

Sediment bulking tends to occur in the surface layer of hydrophobic soils. Hydrophobic soils create a condition where excess water that cannot penetrate deeply saturates the upper few centimetres of soil. This surface material then fails as small-scale debris flows. In addition, water runoff can incorporate so much loose material that sediment concentrations get high enough for the flow to behave as a debris flow. Sediment bulking is probably the most important debris-flow producing process after a fire.

GEOMORPHOLOGICAL HAZARDS are not limited to the first few rainstorms after a fire. Research by Ramon Arrowsmith in the Phoenix, Arizona, region indicates enhanced flash flooding potential decades after a brush fire. Even in forested regions, the supply of loosened material continues to deliver dry ravel sediment, hyperconcentrated flows and debris flows to stream systems for years after a fire.

The link between wildfire and increased erosion leading to large sedimentation events was made as early as 1949 by P.B. Rowe and colleagues working in southern California. They developed the concept of a 'fire–flood sequence' that has been studied extensively in a wide variety of river settings including alpine forests such as Yellowstone (Minshall *et al.* 1998), Mediterranean scrub (Shakesby *et al.* 1993) and even desert ranges (Germanoski and Miller 1995). In Yellowstone, for example, Minshall *et al.* (1998) found extensive RILL development, GULLY formation and MASS MOVEMENTS in burned watersheds during the summer of 1989, when post-fire heavy rains and snowmelt generated widespread 'black water' conditions and increased BEDLOAD and SUSPENDED LOAD. After monitoring Yellowstone streams for a decade after its massive wildfire, Minshall *et al.* (1998) stress that post-fire stream studies can yield misleading insights after only a few years since massive stream reorganization can take place seven to nine years after the fire event.

The study of fire remains associated with soils and sediment, called pedoanthracology, provides important insight into prehistoric geomorphic changes associated with fires. Studies of fire-induced ALLUVIAL FANs, of fire remains within uneroded soils, and diagenesis of organic remains into such forms as vitrinite and inertinite provide geomorphologists with insights into palaeoecological conditions that may have influenced the geomorphic landscape seen today (Siffedine *et al.* 1994).

References

Blackwelder, E. (1927) Fire as an agent in rock weathering, *Journal of Geology* 35, 134–140.

Cannon, S.H., Kirkham, R.M. and Parise, N. (2001) Wildfire-related debris-flow initiation processes, Storm King Mountain, Colorado, *Geomorphology* 39, 171–188.

Dorn, R.I. (2003) Boulder weathering and erosion associated with a wildfire, Sierra Ancha Mountains, Arizona, *Geomorphology*, 55, 155–171.

Germanoski, D. and Miller, J.R. (1995) Geomorphic response to wildfire in an arid watershed, Crow Canyon, Nevada, *Physical Geography* 16, 243–256.

Goudie, A.S., Allison, R.J. and McClaren, S.J. (1992) The relations between modulus of elasticity and temperature in the context of the experimental simulation of rock weathering by fire, *Earth Surface Processes and Landforms* 17, 605–615.

Minshall, G.W., Brock J.T. and Royer T.V. (1998) Stream ecosystem responses to the 1988 wildfires, *Yellowstone Science* 6(3), 15–22.

Morris, S.E. and Moses, T. (1987) Forest-fire and the natural soil-erosion regime in the Colorado Front Range, *Annals of the Association of American Geographers* 77, 245–254.

Shakesby, R., Coelho, C., Ferreira, A., Terry, J. and Walsh, R. (1993) Wildfire impacts on soil-erosion and hydrology in wet Mediterranean forest, Portugal, *International Journal of Wildland Fire* 3, 95–110.

Siffedine, A. *et al.* (1994) The lacustrine organic sedimentation in tropical humid environment (Carajas, eastern Amazonia, Brazil) – relationship with climatic changes during the last 60,000 years BP, *Bulletin de la Société Géologique de France* 165, 613–621.

Swanson, F.J. (1981) Fire and geomorphic processes, in M.A. Mooney *et al.* (eds) *Fire Regimes and Ecosystem Properties*, 401–420, Washington, DC: US Department of Agriculture General Technical Report WO-26.

RONALD I. DORN

FIRST-ORDER STREAM

STREAM ORDERING is based on the premise that stream size is related to the area contributing to runoff. This provides a method of ranking the relative size of streams within a catchment. The term first-order stream originates from ideas initially proposed by R.E. Horton in the 1930s (Horton 1932, 1945). Horton devised a method of classifying links in a stream network using a system of ordering. Under such a scheme the smallest unbranched streams in a catchment are designated first order. The combination of two first-order streams results in a second-order stream and so forth through successively larger links as additional streams join the network (Figure 60a). This original idea soon led to a proliferation of ordering schemes each providing a development

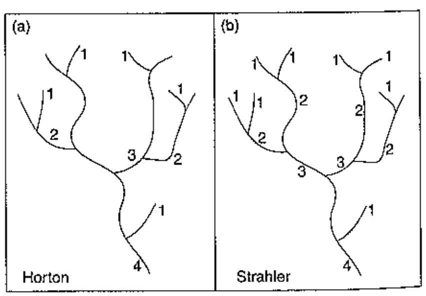


Figure 60 Comparison of stream and segment ordering methods: (a) Horton, (b) Strahler

or refinement of previous ones. Of particular note is the Strahler scheme (Strahler 1952) that begins like the Horton scheme, with the smallest channels being classified as first-order links; but higher order links are only generated when two links of equivalent order are joined (Figure 60b). The highest order generated by this mechanism is often used to classify drainage basins, e.g. a third-order drainage basin. Stream orders vary from the smallest first-order streams to the world's largest rivers that approach twelfth order (Mississippi, Amazon).

The hydrologic response of a stream channel is in part a function of its stream order. Stream order can be used to quantify other aspects of a watershed. These include the Bifurcation Ratio, R_b. The bifurcation ratio (R_b) is defined as the ratio of the number of streams of any order (N_i) to the number of streams of the next highest order. Horton (1945) found that this ratio is relatively constant from one order to another. Values of R_b typically range from the theoretical minimum of 2 to around 6. Typically, the values range from 3 to 5. The bifurcation ratio is calculated as

$$R_b = N_i/N_{i+1}$$

These are important geomorphic parameters that describe the structure and functioning of drainage basins. In the past, calculating these measures was extremely time consuming as catchment boundaries need to be carefully defined. However, these analyses are now routinely undertaken using GIS, which has the potential to provide rapid, accurate and automatic recognition of stream network links (Morris and Heerdegen 1988). This is often based on the topographic definition of streams based on contour crenulation and headwater divide delimitation. In this respect, an advantage of the Strahler scheme is that it retains the same common nomenclature for all similar sized channel links. Thus first-order streams are consistently identified as the smallest channels in a catchment. This is useful because streams with similar attributes, and a similar relative position in the network, are grouped in the same order. Hence, first-order streams tend to have common characteristics. These common characteristics are, however, dependent on the scale at which the channel links are defined, e.g. whether they are mapped from published maps or surveyed in the field. This raises important issues about consistency in definition of network properties (Blyth and Rodda 1973; Mark 1983) and highlights the property that most stream networks are dynamic, so the extent of the network varies in time from storm to storm and across seasons, years and decades.

Topography is not the only criterion used to distinguish first-order channels. First-order streams may also be defined on the basis of flow duration sufficient to sustain aquatic biota year round. In this respect, a first order channel must be by definition permanent, connected to the main stream network and convey runoff from a defined CONTRIBUTING AREA.

The greatest frequency of first-order streams tend to be found in the headwaters of catchments where channels tend to be small, confined, have steeper slopes and individually contribute only small amounts of stream discharge (Wohl 2000). In terms of the overall network, first-order channels defined by a Strahler ordering scheme commonly represent 50–60 per cent of the total stream length in a third-order drainage basin (Strahler 1964). During storms or prolonged wet periods the size of the network will expand and first-order channels may extend up hillslopes as ephemeral water flows are maintained for short periods. The extension of the permanent first-order network beyond the channel head represents a dynamic link. The coupling between the channel head and the network of hillslope hollows upslope usually defines a diffuse topographic network of zero-order basins (Dietrich *et al.* 1987). These zero-order basins are small unchannelled valleys. These form HILLSLOPE HOLLOW networks on slopes which focus runoff and sediment transport via saturated overland flow and gully and debris flows. In general terms, as stream order increases sediment yield per unit area tends to decline as HILLSLOPE-CHANNEL COUPLING becomes less effective.

References

Blyth, K. and Rodda, J.C. (1973) A stream length study, *Water Resources Research* 9, 1,451–1,461.

Dietrich, W.E., Reneau, S.L. and Wilson, C.J. (1987) Overview: zero-order basins and problems of drainage density, sediment transport and hillslope morphology, in *Erosion and Sedimentation in the Pacific Rim*. IAHS Publication 165, 27–37.

Horton, R.E. (1932) Drainage basin characteristics, *Transactions of the American Geophysical Union* 13, 350–361.

——(1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, *Geological Society of America Bulletin* 56, 275–370.

Mark, D.M. (1983) Relations between field-surveyed channel networks and map-based geomorphometric measures, Inez, Kentucky, *Annals of the Association of American Geographers* 73, 358–372.
Morris, D.G. and Heerdegen, R.G. (1988) Automatically derived catchment boundaries and channel networks and their hydrological applications, *Geomorphology* 1, 131–141.
Strahler, A.N. (1952) Hypsometric (area-altitude) analysis of erosional topography, *Geological Society of America Bulletin* 63, 1,117–1,142.
——(1964) Quantitative geomorphology of drainage basins and channel networks, section 4-II, in V.T. Chow (ed.) *Handbook of Applied Hydrology*, 4–39, New York: McGraw-Hill.
Wohl, E. (2000) *Mountain Rivers*, Water Resources Monograph 14, Washington, DC: American Geophysical Union.

SEE ALSO: drainage basin; GIS; runoff generation

JEFF WARBURTON

FISSION TRACK ANALYSIS

Fission track analysis (FTA) is a thermochronometer that provides detailed information on the thermal history of rocks, most usually for temperatures below 350 °C (using zircon) and below 110 °C (using apatite). When a rock has cooled rapidly from its temperature of formation (e.g. a rapidly cooled lava), the technique may provide the age of formation of that rock (hence 'fission track dating') but the technique can be applied in any situation in which low-temperature thermal history is required and the appropriate minerals are present. In geomorphological applications, the technique exploits the increase in temperature with depth in the Earth's crust (the geothermal gradient). This temperature increase means that the low-temperature thermal history of an apatite or a zircon now at the Earth's surface (or in a drill hole) is a record of that mineral's passage through the crust to the sampling point (surface or drill hole). The principal application of FTA in geomorphology is therefore to elucidate the long-term DENUDATION that brings the target mineral(s) to the Earth's surface. For a surface temperature of 20 °C and a geothermal gradient of 25 °C km^{-1}, FTA in apatite provides a denudational history for the upper *c.*4 km of the crust (i.e. below about 110 °C). (URANIUM-THORIUM)/HELIUM ANALYSIS ((U-Th)/He analysis) in apatite provides a shallower denudational history from a lower temperature of *c.*75 °C. In geomorphological studies, in which the final stages of crustal denudation leading to the present topography are of interest, the thermochronometers most often used are the lowest temperature (i.e. shallowest), namely, apatite FTA and (U-Th)/He analysis. If all three low-temperature thermochronometers (i.e. the two in apatite plus zircon FTA) all yield essentially the same ages, then it is clear that denudation (and the associated cooling of the crust through the three thermochronometers' temperature ranges) have occurred very rapidly.

FTA relies on counting the number, and measuring the lengths, of minute damage paths (defects or 'tracks') produced when the heavy daughter products of ^{238}U fission in a mineral's crystal lattice travel away from each other at high speed through the lattice. The tracks are only *c.*5 nm in diameter and are widened slightly by etching in a weak acid during sample preparation, so as to make them visible under microscope. Etched tracks are about 1–2 µm in diameter and up to about 16 µm long. The tracks are produced continuously at a known rate, dependent on the U-content.

In order to reconstruct, from the sample's cooling history, the denudation necessary to bring the sample to the Earth's surface, a knowledge of the geothermal gradient at the time the sample was exhumed is necessary. This geothermal gradient provides the crustal depths associated with the temperatures from which the sample was exhumed. The 'ancient' geothermal gradient is usually unknown and an 'appropriate' geothermal gradient is often assumed based on likely modern analogues of the tectonic and thermal setting of the sample locality at the time of exhumation. If a vertical profile of FT samples is available (for example, from a drill hole or from a mountain side), the gradient of the elevation–age profile provides the geothermal gradient.

In simple terms, the number of tracks is a function of the time since the sample cooled sufficiently for the tracks to be retained (i.e. cooled below about 110 °C in apatite), and the U-content of the mineral in the areas of the grain in which the tracks have been counted and their lengths measured. The fission track age is derived by the application of the standard radiometric dating formula but with the amount of decay product ('daughter') of the dating technique's radioactive decay system replaced by the number of tracks.

Lower and/or more variable rates of denudation through time result in more complex cooling histories, which can be elucidated using frequency distributions of track lengths (the 'track length distribution'). Tracks form continuously as a result of ^{238}U fission but in apatite, for example, they are annealed (repaired) geologically instantaneously above a temperature of about 110 °C. Below 110 °C, apatite fission tracks are only partially annealed and are increasingly retained at temperatures down to surface temperature. This temperature range in which tracks are partially annealed (repaired) is the partial annealing zone (PAZ), and there is a range of views as to the effective lower limit of the PAZ. Strictly, fission tracks may be annealed even at room temperature but some authors set the effective lower boundary of the PAZ at *c.*60 °C. Track annealing is by repair at the ends, resulting in shorter tracks. The duration of the sample's residence in the PAZ is therefore reflected in the frequency distribution of track lengths, shorter track lengths reflecting longer residence time in the PAZ.

Statistical temperature–time paths can be calculated to match the measured fission track age and track length distribution, giving a complete description of thermal history of the apatite below 110 °C, and hence of the sample's trajectory to the surface as a result of denudation. Figure 61 shows the ways in which different fission track ages and track length distributions reflect different cooling histories. In A, the sample cooled very rapidly at 100 Ma ago, and the fission track age (99.8 Ma; the upper number of the three within the plot) is essentially the same as the age of the cooling event. The rapid cooling is reflected in the long mean track length (15 µm; the middle number in the plot) and the very low standard deviation of the track length distribution (1.07 µm; the third number). The track length data have a high, unimodal, narrow distribution in the histogram of the track length distribution.

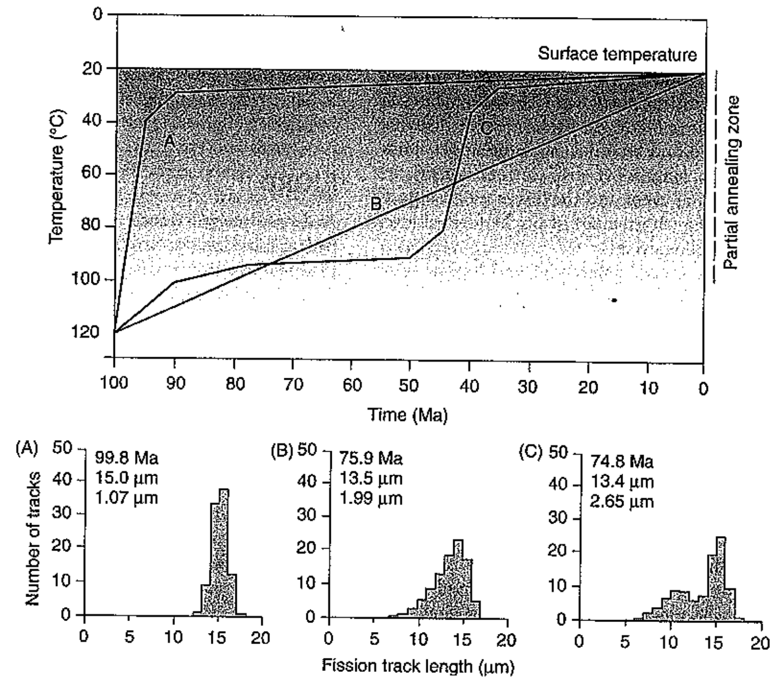


Figure 61 Fission track ages and track length in relation to cooling histories (based on Gleadow and Brown 2000: figure 4.3)

Sample A's very rapid cooling could be the result of very high rates of subaerial denudation (probably combined with ongoing rapid uplift to drive the denudational exhumation) or it could be associated with tectonic denudation in which very high rates of uplift lead to detachment of slabs of crust which slide away by gravity along decollements, thereby cooling the underlying crust. Sample B has experienced steady cooling from 100 Ma ago to the present. The average track length is shorter (13.5 μm) and the distribution broader (s = 1.99 μm), both measures being reflected in the broader histogram with a longer 'tail' into the short track lengths (reflecting the greater time that tracks have spent in the PAZ after formation). Note how B's fission track age (75.9 Ma) bears no obvious relation to the cooling event or to any particular depth in the crust for determining a rate of denudation. The determination of the rate of denudation requires modelling of the cooling history of the sample (in effect determining the cooling history, as in the upper diagram, from the age and track length distribution in the lower diagram). In the more complex cooling history of C (two discrete cooling events: one between 100 and 90 Ma and the second at about 45 Ma), the fission track age (74.8 Ma) relates to neither cooling event. The track length distribution is broad and bimodal, with the upper mode (long track lengths of c.15 μm) reflecting the 45 Ma cooling event and the lower mode reflecting annealing (shortening) of tracks formed after the first cooling event.

There are several inferential and logical steps involved in converting FTA data to a geomorphological history and an amount of denudation (e.g. Gleadow and Brown, 2000). Notwithstanding the uncertainties and assumptions associated with these steps, FTA has been successfully applied to elucidate long-term landscape development in a range of settings. Application in active orogenic settings of FTA in conjunction with higher temperature thermochronometers, such as the $^{40}Ar/^{39}Ar$ system, and lower temperature systems, such as *(uranium-thorium)/helium analysis* in apatite, has very convincingly demonstrated that denudation of these settings is very rapid. When various thermochronometric systems yield the same rates of denudation, it is argued that there is a dynamic equilibrium between denudation and the ongoing tectonic uplift necessary to drive the flux of crust through the Earth's surface where it is removed by denudation at the same rate as uplift. The processes and sequences of events associated with lithospheric extension and subsequent PASSIVE MARGIN development have also been widely elucidated using FTA. FTA data along many passive continental margins, especially the data from closest to the new margin, exhibit rapid cooling events (long, unimodal track length distributions) at about the time of break-up. These FTA data are interpreted in terms of rapid denudation of the nascent or new continental margin at about the time of breakup, in response to one or more of the following: thermally driven active or passive uplift and denudation of the pre-breakup rift shoulders; rapid denudation of the new margin in response to the new BASE LEVEL for denudation that is provided by the formation of a new ocean basin adjacent to the margin; and ongoing flexural isostatic uplift of the new margin in response to this accelerated denudation.

Reference

Gleadow, A.J.W. and Brown, R.W. (2000) Fission-track thermochronology and the long-term denudational response to tectonics, in M.A. Summerfield (ed.) *Geomorphology and Global Tectonics*, 57–75, Chichester: Wiley.

PAUL BISHOP

FJORD

A deeply incised trench or trough excavated in bedrock by long-term glacial erosion and occupied by the sea during periods of glacier recession. Spectacular fjordic scenery occurs along the coasts of British Columbia in Canada, Alaska, southern Chile, Greenland, northern and eastern Iceland, Spitsbergen, Fiordland in New Zealand, the Canadian arctic islands and western Scotland. The longest fjords are Nordvestfjord/Scoresby Sund in Greenland (300 km), Sognefjord in Norway (220 km) and Greely Fjord/Nansen Sound in the Canadian arctic (400 km).

Troughs and fjords have distinctive cross and long profiles, referred to as U-shaped but best approximated by the formula for a parabola:

$$V_d = aw^b$$

where w is the valley half width, V_d is valley depth and a and b are constants. However, true cross profiles deviate from this mathematical parabola largely due to the production of breaks in slope by pulsed erosion through time. These effects have been modelled by Harbor *et al.* (1988) by imparting a valley glacier on a fluvial, V-shaped valley. Basal velocities below a glacier are highest part way up the valley sides and lowest below the glacier margins and centre line. By assuming that the erosion rates are proportional to the sliding velocity, the greatest erosion occurs on the valley sides, thereby causing broadening and steepening of the valley. The development of the steep sides of troughs and fjords is aided by PRESSURE RELEASE or dilatation in the bedrock. This is the development of fractures parallel to the ground surface. Such fractures weaken rock masses, thereby facilitating subsequent subglacial erosion. Dilatation is most likely to take place immediately after deglaciation when the glacier overburden has been removed and freshly eroded rock surfaces are exposed.

Overdeepenings along fjord and trough long profiles separated by sills or thresholds appear to represent areas of increased glacier discharge such as at the junctions of tributary valleys or where fjord narrowing occurs. The area of deepest erosion in a fjord marks the location of the long-term average position of maximum glacier discharge. Fjord mouths are often characterized by STRANDFLATS, likely due to the fact that the erosion capacity of the outlet glaciers is severely reduced due to glacier buoyancy and eventual ice flotation in the sea in addition to the flow divergence induced by the more open topography.

The planform of many fjords clearly reveals fluvial or structural origins. For example, the sinuous forms and dendritic patterns of some fjords suggest that they are glacially overdeepened preglacial fluvial valleys and rectilinear fjord networks have been linked to large-scale structural features such as faults and grabens. Moreover, the close association between linear fjord alignments and intersecting lines of regional fracture have led to purely tectonic theories for fjord initiation. The survivial of preglacial landforms and sediments on upland areas between fjords demonstrates that the deep glacial incision is selective, hence the use of the term *selective linear erosion* to describe the development of fjord and trough systems. It is most likely that pre-existing valley systems, whether fluvial and/or tectonic in origin, will contain thicker ice during glaciations and therefore act as major conduits for glacier flow from the centres of ice dispersal, especially if they are oriented parallel to regional glacier flow. Greater ice thicknesses and concomitant preferential ice flow down such valleys will result in greater frictional heat, increased pressure melting and widespread basal sliding. Conversely, on the plateaux between fjords, cold-based ice will dominate and protect underlying preglacial features from glacial erosion. The occurrence of a preglacial land surface on the plateaux surrounding Sognefjord has allowed the calculation of a fjord erosion rate (Nesje *et al.* 1992). Approximately 7,610 km^3 of material has been removed from the fjord by glacial erosion, yielding erosion rates ranging from 102 to 330 cm kyr^{-1} depending upon the amount of time that glaciations have dominated the region.

The dimensions of fjords appear to be scaled to the amount of ice that was discharged through them, several researchers having demonstrated that relationships exist between fjord size and glacier contributing area. The strength of these relationships also varies between regions. For example, Augustinus (1992) demonstrated that fjords in British Columbia are 2.5 times deeper and 2.4 times longer than New Zealand fjords even though the contributing areas are comparable in size. This suggests that glacial erosion is more intense in British Columbia probably due to the fact that water depths are shallower offshore than in New Zealand and are therefore less capable of floating the fjord glaciers. In addition, the lengths of the former British Columbia fjord glaciers were much greater than those of the New Zealand palaeo-glaciers, the former having been nourished by an inland ice sheet.

References

Augustinus, P.C. (1992) Outlet glacier trough size–drainage area relationships, Fjordland, New Zealand, *Geomorphology* 4, 347–361.

Harbor, J.M., Hallet, B. and Raymond, C.F. (1988) A numerical model of landform development by glacial erosion, *Nature* 333, 347–349.

Nesje, A., Dahl, S.O., Valen, V. and Ovstedal, J. (1992) Quaternary erosion in the Sognefjord drainage basin, western Norway, *Geomorphology* 5, 511–520.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, 350–356, 362–366, London: Arnold.

England, J. (1987) Glaciation and the evolution of the Canadian high arctic landscape, *Geology* 15, 419–424.

Harbor, J.M. (1992) Numerical modelling of the development of U-shaped valleys by glacial erosion, *Geological Society of America Bulletin* 104, 1,364–1,375.

Holtedahl, H. (1967) Notes on the formation of fjords and fjord valleys, *Geografiska Annaler* 49A, 188–203.

Løken, O.H. and Hodgson, D.A. (1971) On the submarine geomorphology along the east coast of Baffin Island, *Canadian Journal of Earth Sciences* 8, 185–195.
Nesje, A. and Whillans, I.M. (1994) Erosion of Sognefjord, Norway, *Geomorphology* 9, 33–45.
Roberts, M.C. and Rood, R.M. (1984) The role of ice contributing area in the morphology of transverse fjords, British Columbia, *Geografiska Annaler* 66A, 381–393.
Shoemaker, E.M. (1986) The formation of fjord thresholds, *Journal of Glaciology* 32, 65–71.

DAVID J.A. EVANS

FLASH FLOOD

Flash flood denotes an abrupt rise in the discharge of a river or stream, providing an event of short duration. The term has conventionally been associated with ephemeral flow regimes in which the majority of events are rain-fed. The flood is discrete. It impinges on a channel bed that is initially dry and is exhausted within a short interval – a few hours in the case of small drainage basins or a few days where basin size involves longer travel time. Because of this, flash floods are commonly associated with deserts and semi-deserts of low to middle latitudes, flow in high-latitude deserts resulting rather from the slow release of water in the form of snowmelt – a seasonal freshet lasting continuously for weeks or months. However, the term has been used more widely to describe a sudden significant increase in discharge where the annual flow regime is intermittent or even perennial. In environments such as those with a Mediterranean-type climate, flow dwindles seasonally so that flood runoff in summer may add dramatically to a pre-existing trickle. In these circumstances, the perception of an observer will be that the rising limb of the flood hydrograph is steep, occupying tens of minutes rather than hours, a pattern which contrasts with that more typical of runoff during the wet season. To warrant the descriptor, the magnitude of the flood peak will also have been remarkable in causing nuisance or damage, or, indeed, loss of human life. It is debatable whether, here, the flood hydrograph should be described as 'flashy' or 'flashier' in relation to the norm rather than given the epithet 'flash flood'.

In desert or semi-desert environments, where dry channel conditions are the norm, (see WADI) flash floods are usually remarkable regardless of magnitude. Here, in contrast with Mediterranean or temperate environments where events of this type are a feature of the 'dry season', floods are more often than not associated with incursions of monsoonal airmass (as in the Sonoran Desert and the Saudi Arabian peninsula) or with the regular latitudinal shift of the Inter-Tropical Convergence Zone (as in East Africa). In such areas, these are events of the rain season(s), but rainfall is extremely uncertain, making events even more memorable if, by chance, they are witnessed. Although widespread, low intensity rainfall can generate runoff in these areas, flash floods are more likely to be the product of wandering cellular convective storms. The wetted 'trail' or footprint of these is usually only a few kilometres across. Atmospheric dynamics dictate that such storm systems that are capable of releasing rain of sufficient magnitude and intensity will be separated by several tens to several hundreds of kilometres. The likelihood that a drainage basin will receive sufficient rainfall to generate runoff (see RUNOFF GENERATION) depends upon its position in relation to the trajectory of each storm. Small basins (in the order of 10^1 km^2) may experience an event only infrequently, perhaps staying dry even though floods occur in the vicinity. In this case, it may be that a flood series for one basin is developed from events in years that do not contribute to the series of a neighbouring stream. In basins of moderate size (several 10^2 km^2), the storm cell is frequently smaller than the basin. Indeed, in this case, the flood may move down-channel into parts of the basin that have not experienced rain. This is a circumstance that provides the greatest danger for the unwary and is not uncommonly the cause of human mortality, especially where the dry river bed has provided an apparently convenient location for overnight encampment.

The significance of the variable spatial coincidence of storm and drainage basin is that the flood hydrograph can take on a variety of shapes. This is, in part, because different sub-basins may contribute to each event and storms may move up or down catchment, depending on local atmospheric dynamics, so affecting the gathering time of contributions from each tributary. This means that it is more difficult to define a typical flood hydrograph (as in, e.g. unit hydrograph analysis) in a desert or semi-desert setting, not only because the frequency of events is low but also because each runoff event may possess unique characteristics. There is, however, evidence from one drainage basin in the Asir Escarpment of Saudi Arabia that flood volume can be approximated from a parameter such as flood peak discharge, and finite element models of rainfall runoff have been developed with some success for predicting flood waves in small basins in Oman and Arizona. Despite these, the relations between runoff and storm characteristics such as rainfall amount and intensity are often chaotic so that predictability of event frequency and magnitude is low even if rainfall is being monitored by spatially inclusive means such as radar. In all but rare instances where research catchments have been established, rain gauge density and disposition will be either inadequate or, more often, non-existent.

Flash floods are undoubtedly dramatic, if only because of the stark contrast between the event itself and the much longer intervening period when the channel is dry. In southern Israel, at the eastern edge of the Sahara's hyper-arid core, long-term monitoring has revealed that the frequency of events is, on average, much less than one a year, but there can be periods of several years when no runoff occurs. In the semi-arid northern Negev, with a rainfall of *c.*200–300 mm per year, the number of events that occurs in moderate-sized basins ranges from zero to seven. Here, on average, an ephemeral channel is occupied by flow for about 2 per cent of the year, or about seven days.

Perhaps the most dramatic aspect of flash floods is the arrival of a bore. This may be the first that the observer is aware of rainfall, which may have occurred well up-catchment. Field monitoring in semi-arid areas, where vegetation may be sparse at the start of the rain season, has shown that time to ponding is short – typically in the order of a few minutes, depending on the infiltration capacity of the local soils – even under modest rainfall intensities. An observer caught out in the rain undergoes a curious sensation that the ground is moving as a glistening sheet of OVERLAND FLOW slips towards the channel system. Here, high drainage densities, developed in response to the easy and quick shedding of water, ensure rapid concentration of flow and the birth of a flash flood.

A flash flood bore takes on a number of forms. The rapidly advancing 'wall of water' is almost certainly a figment of imagination encouraged by the panic of moving to a place of safety. Indeed, the type of bore most commonly caught on camera is comparatively shallow, with low trailing water-surface slope. However, a few examples have been photographed where the bore reaches a height of about half a metre (Plate 49). Of those few measurements that have been made of bore advance, velocities range from 0.5 to 2 m s^{-1}, the rate depending directly on bore height. This is equivalent to a stiff walking pace for a human being and one might wonder, therefore, what reasons there are for the number of fatalities that are reported. The problem for those unfortunate to be caught napping is that, following the passage of the bore, water levels rapidly increase. One fully documented example has indicated an average rise of a quarter of a metre per minute, so that the water surface was at waist height within two and head height within ten minutes of the start of hydrograph rise. By this time, average flow velocity is in excess of 3 m s^{-1} and increasing to values greater than 5 m s^{-1} (Figure 62).

Plate 49 Flash flood bore in Nahal Eshtemoa, northern Negev, advancing over dry bed at about 2 m s^{-1}. Note that the immediate area has had no rain

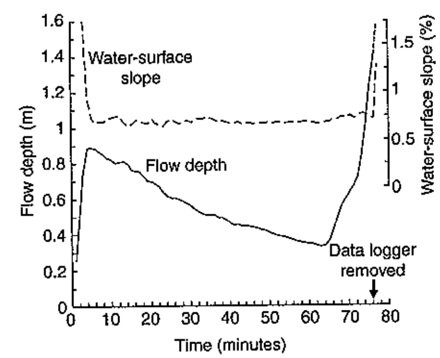


Figure 62 Hydrograph and water surface slope of the flash flood on Nahal Eshtemoa shown in Plate 49

However, a flash flood bore is not stealthy. If it advances over a gravel bed, the cacophony of grains being thrown against each other can be heard several hundred metres away. Flotsam is also characteristic of flash flood bores, the infrequent flow sweeping up LARGE WOODY DEBRIS and other organic matter that has fallen into the channel between events and adding to the general confusion that is already inherent. Indeed, some have reported hearing the clash of tree trunks, etc. several kilometres ahead of the bore's arrival.

Although the bore of a flash flood is its most spectacular feature, another unique but hidden characteristic is the loss of a significant fraction of flow to the dry bed. These are dubbed transmission losses. They are determined in part by the magnitude of the flow and hence the wetted perimeter. For a small heavily gauged ephemeral channel in Arizona, examples show that transmission losses to the bed in each kilometre of channel can account for as much as 6 per cent of the flow. This points to another important characteristic of flash floods in desert and semi-desert settings – many fail to reach the terminal ALLUVIAL FAN.

Further reading

Bull, L.J. and Kirkby, M.J. (eds) (2002) *Dryland Rivers: Processes and Management in Mediterranean Climates*, Chichester: Wiley.

Reid, I., Laronne, J.B., Powell, D.M. and Garcia, C. (1994) Flash floods in desert rivers: studying the unexpected, *EOS, Transactions American Geophysical Union* 75, 452.

Reid, I., Laronne, J.B. and Powell, D.M. (1998) Flashflood and bedload dynamics of desert gravel-bed streams, *Hydrological Processes* 12, 543–557.

IAN REID

FLAT IRON

Term used to designate relic slopes whose morphology resembles a reversed iron. They are also known as talus flat irons or triangular slope facets and develop at the foot of scarps in mesas and cuestas. The slope deposits, which locally contain datable charcoal, ashes or pottery remains, may grade in the distal sector to cover pediments or fluvial or lacustrine terraces.

The most widely accepted genetic model relates the development of talus flat irons to climatic changes. The accumulation processes in the slopes prevail during humid periods whereas the reduction in the vegetation cover during dry periods favours rilling and gullying processes. Successive climate changes give place to different generations of relict slopes whose relative chronology can be inferred from their spatial distribution. Up to five generations of flat irons have been identified in the three main Tertiary basins of Spain. The slope deposits of the flat irons dated with ^{14}C correspond to cold periods. The youngest facet generation fits with Upper Holocene Neoglaciation episodes and the two previous generations correlate to Heinrich events (H_3 and H_4) that indicate cold periods (Gutierrez *et al.* 1998).

Reference

Gutierrez, M., Sancho, C., Arauzo, T. and Peña, J.L. (1998) Evolution and paleoclimatic meaning of the talus flat irons in the Ebro Basin, northeast Spain, in A.S. Alsharham, K.W. Glennie, G.L. Whittle and C.G.St.C. Kendall (eds) *Quaternary Deserts and Climatic Change*, 593–599, Rotterdam: Balkema.

SEE ALSO: slope, evolution

M. GUTIERREZ-ELORZA

FLOOD

A flood is a flow of water greater than the average flow along a river. A flood may be described in terms of its magnitude. On a given river, for example, any discharge exceeding $1,000\,m^3/s$ might be designated a flood. A flood may also be described by its recurrence interval; a 100-yr flood occurs on average once every 100 years. Or a flood may be described as any flow that overtops the banks or LEVEES along a channel and spreads across the FLOODPLAIN.

Floods along inland rivers result from precipitation or from a damburst. When water rapidly flows downslope from snowmelt, rain-on-snow, or various types of rainfall, the baseflow from subsurface water that keeps some stream channels flowing during dry periods is augmented by runoff (see RUNOFF GENERATION). As discharge increases in the channel during the rising limb of the flood, the channel boundaries may be eroded, and both SUSPENDED LOAD and BEDLOAD sediment transport are likely to increase. Once the input of runoff to the channel declines, sediment transport is likely to decrease and sediment may be deposited along the channel during the falling limb of the flood. Floodwaters in a channel commonly rise more rapidly than they fall during all types of floods, but this

difference is most pronounced for damburst floods. Dams built by humans and naturally occurring LANDSLIDES, ice jams, glacial moraines, glacial ice dams, or beaver DAMS may fail suddenly, prompting catastrophic drainage of the water ponded behind the dam. Along with FLASH FLOODS, damburst floods are often the most unexpected and damaging floods. Damburst floods may have a peak discharge more than an order of magnitude larger than the peak discharges created by meteorological floods along a river. This large discharge may generate high values of STREAM POWER that cause substantial erosion and deposition along the flood path. OUTBURST FLOODS generated by the failure of natural dams ponding meltwater from the great continental ice sheets during the Pleistocene shaped such dramatic landscapes as the Channeled SCABLAND of the northwestern United States.

Floods along coastal rivers may also result from STORM SURGES, TSUNAMIS, or other anomalously large waves or tides backflooding upstream from the ocean. Low-relief coastal areas such as those found in Bangladesh may be particularly susceptible to such floods.

The largest measured historical floods generated by precipitation have occurred primarily between 40°N and 40°S latitude, usually near coastal areas where the onshore movement of warm, moist airmasses into the continental interior produces intense and widespread precipitation (Costa 1987). The envelope curve of maximum rainfall-runoff floods is mathematically described by $Q = 90A$ for drainage areas less than 100 square kilometres, and $Q = 850A^{0.357}$ for larger drainage areas, where Q is peak discharge in cubic metres per second and A is drainage area in square kilometres (Herschy 1998).

The importance of a flood relative to smaller flows in shaping channel and valley morphology will depend on the magnitude and duration of the hydraulic forces generated during the flood in comparison to the erosional resistance of the channel boundaries, and on the recurrence interval of the flood. A channel formed on bedrock or very coarse alluvium may have such high boundary resistance that only a flood generates sufficient force to erode the channel boundaries. This effect may be enhanced where a deep, narrow channel and valley geometry concentrate floodwaters such that flow depth increases rapidly with discharge, giving rise to high stream power. In contrast, a channel bordered by a broad floodplain will have much less increase in flow depth with increasing discharge, and the flood may not have a substantially greater capacity for erosion and sediment transport than do smaller flows along the channel. Channels in which geometry and sediment transport reflect primarily floods are likely to have a flashy hydrograph, abundant coarse sediment load, a high channel gradient, highly turbulent flow, and shifting, erodible banks (Kochel 1988). Geomorphic change during floods is likely above a minimum threshold (see THRESHOLD, GEOMORPHIC) of approximately $300\,W\,m^{-2}$ of unit stream power for alluvial channels (Magilligan 1992). The threshold for bedrock channels may be expressed as $y = 21x^{0.36}$, where y is stream power per unit area and x is drainage area (Wohl *et al.* 2001). In steep channels with abundant sediment, flows may alternate downstream among water-floods, DEBRIS FLOWS and HYPERCONCENTRATED FLOWS.

Measures to reduce hazards to humans associated with floods date back several millennia. Such measures include impoundments to regulate water flow; channelization to increase the flood conveyance of channels; levees to confine floodwaters; warning systems to help alert and evacuate humans at risk; and engineering designs which reduce flood damage to structures. Despite this long history of river engineering and flood mitigation, property damage from floods continues to increase worldwide as population density and building in flood-prone areas increase, and as land uses across drainage basins alter runoff generation. Along rivers where alteration of the natural flow regime has reduced or eliminated floods, aquatic and riparian species adapted to flooding have declined in extent and diversity. River rehabilitation and restoration measures are now being applied to some of these rivers in an attempt to mitigate damages caused by the absence of floods.

References

Costa, J.E. (1987) A comparison of the largest rainfall-runoff floods in the United States with those of the Peoples Republic of China and the world, *Journal of Hydrology* 96, 101–115.

Herschy, R.W. (1998) Floods: largest in the USA, China and the world, in R.W. Herschy and R.W. Fairbridge (eds) *Encyclopedia of Hydrology and Water Resources*, 298–300, Dordrecht: Kluwer Academic.

Kochel, R.C. (1988) Geomorphic impact of large floods: review and new perspectives on magnitude and frequency, in V.R. Baker, R.C. Kochel and P.C. Patton (eds) *Flood Geomorphology*, 169–187, New York: Wiley.

Magilligan, F.J. (1992) Thresholds and the spatial variability of flood power during extreme floods, *Geomorphology* 5, 373–390.
Wohl, E., Cenderelli, D. and Mejia-Navarro, M. (2001) Channel change from extreme floods in canyon rivers, in D.J. Anthony, M.D. Harvey, J.B. Laronne and M.P. Mosley (eds) *Applying Geomorphology to Environmental Management*, 149–174, Highlands Ranch, CO: Water Resources Publications.

SEE ALSO: bankfull discharge; floodout; palaeoflood; sediment rating curve

ELLEN E. WOHL

FLOODOUT

A floodout is a site at the downstream end of a river where channelized flow ceases and floodwaters spill across adjacent, unchannelled, alluvial surfaces. The term has been most widely used in connection with ephemeral channels in arid central Australia (Tooth 1999a) but it has also been applied to discontinuous gullies (see GULLY), intermittent channels and perennial channels in semi-arid, subhumid and humid regions of eastern Australia and southern Africa (Fryirs and Brierley 1998; Tooth *et al.* 2002).

Floodouts form as a result of various factors including downstream decreases in discharge, downstream decreases in gradient, and aeolian or bedrock barriers to flow (Tooth 1999a). These factors commonly act in combination. For example, along many arid or semi-arid rivers, discharge decreases downstream owing to factors such as infiltration into normally dry channel beds, evaporation, hydrograph attenuation and a lack of tributary inflows. Gradient also commonly decreases owing to channel-bed AGGRADATION or lithological/structural factors, such as a change from a harder to a weaker lithology underlying the channel bed. In combination, these discharge and gradient decreases mean that unit STREAM POWER and sediment transport capacity also decrease, which in turn leads to a downstream reduction in the size of the channel and diversion of an increasing proportion of floodwaters overbank. This is often exacerbated by the presence of aeolian or bedrock barriers, such as longitudinal dunes (see DUNE, AEOLIAN) that have formed across the river course. Eventually, the channel loses definition and disappears entirely, and the remaining floodwaters spill across the floodout as a sheet flow (see SHEET EROSION, SHEET FLOW, SHEET WASH). This process is often referred to as 'flooding out' but strictly speaking the term 'floodout' and its derivatives should be used for the fluvial form only.

Floodouts can form in river catchments of widely different scale and thus the areas of floodouts vary considerably (c.1–1,000 km^2). The location and shape of floodouts, however, are often strongly influenced by local physiography. In central Australia, for instance, floodouts in the northern Simpson Desert are narrow (<500 m) features where rivers terminate between longitudinal dunes and occasional bedrock outcrops but on the relatively unconfined Northern Plains they can reach up to several kilometres wide (Tooth 1999a,b).

'Floodout zone' is a related but broader term that encompasses both the lower reaches of the channel and the floodout itself. Geomorphological and sedimentary features commonly associated with floodout zones include distributary channels, splays, waterholes, PANS, PALAEOCHANNELS and various fluvial-aeolian interactions (Tooth 1999a,b). In addition, two basic types of floodout can be distinguished (Tooth 1999a): (1) terminal floodouts, where floodwaters spill across the unchannelled surfaces and eventually dissipate through infiltration or evaporation; and (2) intermediate floodouts, where floodwaters persist across the unchannelled surfaces and ultimately concentrate into small 'reforming channels'. Reforming channels commonly develop where the unchannelled floodwaters become constricted by aeolian deposits or bedrock outcrops, or where small tributaries provide additional inflow, and they either join a larger river or decrease in size downstream before disappearing in another floodout (Tooth 1999a; Tooth *et al.* 2002).

Formation of a floodout is just one possible end result of the broader processes of channel 'breakdown', 'failure' or 'termination' that can also occur where channels disappear in playas, in permanent wetlands, or on the surfaces of ALLUVIAL FANS. Floodouts, however, are predominantly alluvial features which are normally dry except after flood events or heavy local rains, and thus they differ from saline playas or organic-rich, saturated wetlands. Furthermore, the relatively low gradients (<0.002) and fine-grained deposits typical of floodout zones distinguish them from alluvial fans. Floodout zones have many geomorphological and sedimentological similiarities with 'terminal fans', a term that has been applied to the distal reaches of some inland arid and semi-arid river systems where numerous distributary channels decrease in size downstream and grade into unchannelled, fan-shaped deposits (Mukerji 1976; Kelly and Olsen 1993). Downstream of intermediate floodouts, however, channels can reform, thus showing that floodouts are not necessarily terminal and, as floodouts are often confined laterally by aeolian deposits or bedrock outcrop, neither does alluvial deposition necessarily adopt a fan-shaped form (Tooth 1999a). As such, application of the term 'terminal fan' is inappropriate for many floodouts or floodout zones. In floodout zones, the disappearance of channelized flow means that FLOODPLAINS grade downstream into floodouts. Although the absence of channels makes it difficult to include floodouts within conventional definitions of 'floodplain' or existing floodplain classifications, nevertheless they can be regarded as part of a continuum of floodplain types (Tooth 1999b).

References

Fryirs, K. and Brierley, G.J. (1998) The character and age structure of valley fills in upper Wolumla Creek catchment, South Coast, New South Wales, Australia, *Earth Surface Processes and Landforms* 23, 271–287.
Kelly, S.B. and Olsen, H. (1993) Terminal fans – a review with reference to Devonian examples, in C.R. Fielding (ed.) Current Research in Fluvial Sedimentology, *Sedimentary Geology* 85, 339–374.
Mukerji, A.B. (1976) Terminal fans of inland streams in Sutlej-Yamuna Plain, India, *Zeitschrift für Geomorphologie NF* 20, 190–204.
Tooth, S. (1999a) Floodouts in central Australia, in A. Miller and A. Gupta (eds) *Varieties of Fluvial Form*, 219–247, Chichester: Wiley.
——(1999b) Downstream changes in floodplain character on the Northern Plains of arid central Australia, in N.D. Smith and J. Rogers (eds) *Fluvial Sedimentology VI*, International Association of Sedimentologists, Special Publication 28, 93–112, Oxford: Blackwell Scientific Publications.
Tooth, S., McCarthy, T.S., Hancox, P.J., Brandt, D., Buckley, K., Nortje, E. and McQuade, S. (2002) The geomorphology of the Nyl River and floodplain in the semi-arid Northern Province, South Africa, *South African Geographical Journal* 84(2), 226–237.

Further reading

Bourke, M.C. and Pickup, G. (1999) Fluvial form variability in arid central Australia, in A. Miller and A. Gupta (eds) *Varieties of Fluvial Form*, 249–271, Chichester: Wiley.
Gore, D.B., Brierley, G.J., Pickard, J. and Jansen, J.D. (2000) Anatomy of a floodout in semi-arid eastern Australia, *Zeitschrift für Geomorphologie Supplementband* 122, 113–139.
Mabbutt, J.A. (1977) *Desert Landforms*, Canberra: ANU Press.

SEE ALSO: alluvium; bankfull discharge; flood; hydraulic geometry

STEPHEN TOOTH

FLOODPLAIN

The floodplain is generally considered to be the relatively flat area of land that stretches from the banks of the parent stream to the base of the valley walls and over which water from the parent stream flows at times of high discharge. The sediment that comprises the floodplain is mainly ALLUVIUM derived from the parent stream with minor contributions from aeolian sediment or colluvium from the valley walls. During floods the channel width is increased to include some or all of the floodplain in order to accommodate the increased discharge with relatively smaller increases in velocity and depth than would be the case if the flood discharge were artificially confined within the channel. However, defining the extent of a floodplain at a locality in terms of the area inundated in floods of particular return periods poses problems, since flooding frequency may be a restricting factor. This can be especially problematic in arid and semi-arid areas.

It has been suggested (Wolman and Leopold 1957) that the active floodplain is the area subject to the annual flood (i.e. the highest discharge each year), though this can really only apply to rivers in humid regions. In reality, the active floodplain only forms part of the topographic floodplain, which encompasses the whole valley floor and includes parts of relict floodplains in the form of river terraces (see TERRACE, RIVER) (Plate 50). If the floodplain is defined in terms of the processes (including superfloods) that give rise to it, then the term polygenetic floodplain would apply to most since they result from changes in flow-regime and sediment supply over at least the recent geological past. Nanson and Croke (1992) have proposed the term genetic floodplain, which applies to a generally horizontally bedded landform built from alluvium derived from the present flow-regime of the adjacent stream. This does not take into account the geomorphic history of a floodplain and the processes that have influenced its construction over time, however. A floodplain is a functional part of the whole stream system and forms as

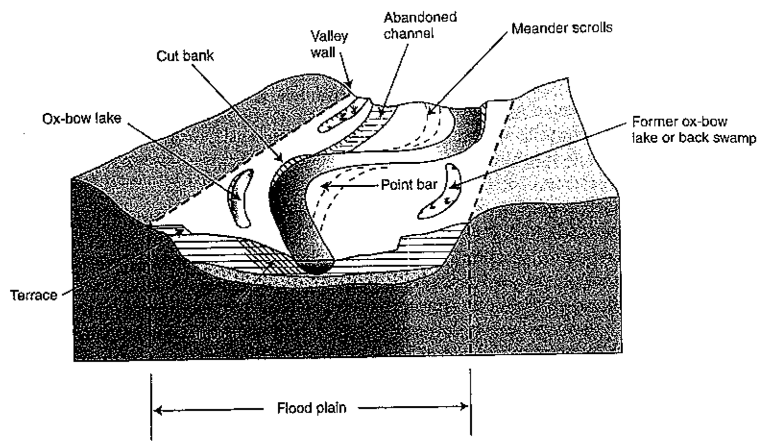


Plate 50 Features of floodplain topography

a byproduct of interrelated processes that, over time, give rise to variable flows and sediment loads derived from the drainage basin.

Floodplain formation

Floodplains are formed by processes that are active both within the channel of the parent stream and during overbank flow. The processes involved are lateral accretion, which takes place within the channel from the formation of bars by movement of relatively coarse bedload; and vertical accretion, which occurs on the floodplain surface due to deposition of finer material from the suspended load during overbank flow. The relative importance of vertical accretion in the formation of floodplains was considered negligible in comparison with within-channel processes (Wolman and Leopold 1957) though it has now been shown that overbank sedimentation can contribute significantly. For example, Nanson and Young (1981) have described the floodplain deposits of streams in New South Wales that have parts of their floodplains dominated by extremely cohesive overbank deposits that prevent the stream from migrating. In lowland rivers in the United Kingdom similar deposits have been described for the Severn (Brown 1987) and the Thames (Lewin 1984) where thick muddy deposits overlie sandy gravels.

Lateral accretion deposits are built up in the channel either as marginal bars that may form in an alternating sequence along relatively straight channels or as point bars that develop on the inside bends of meanders. If the channel migrates laterally by BANK EROSION on one side, the channel dimensions are maintained by compensating deposition on the other bank. The marginal or point bars grow laterally towards the direction of migration and increase in height, with sediment being deposited on them in low-angled, sloping layers overlain by finer material deposited at bankfull and flood stage flow. The sediments within the bars tend to fine upwards as they initially form from the stream's bedload sediment carried by secondary currents from the outer bank region towards the inner bank (Markham and Thorne 1992). Over time, in this way, a stream may rework the entire floodplain sediment as it migrates from one valley side to the other, leaving behind cutoff meanders as oxbow lakes or swamps and traces of the old meander paths as meander scrolls (Plate 50).

Vertically accreted sediment is added to the floodplain surface during overbank flow. As the water in the channel overflows the banks onto the floodplain surface, the flow velocity is reduced as the width of the channel is effectively increased by inclusion of the floodplain. This reduction, in conjunction with an increase in surface roughness if the floodplain is vegetated, causes suspended sediment to be deposited. As the flow of water on the floodplain is slower and shallower than that in the channel, there is a zone of turbulence near to the channel bank which results in a net transfer of momentum and sediment from the channel to the

floodplain. The width of the turbulent zone depends on the relative difference in depths between the channel and floodplain flows but influences the nature of sediment deposition near to the channel. The sediment deposited near to the channel bank tends to be coarser and thicker than that deposited further onto the floodplain as transport competence declines with distance from the channel and away from the influence of the turbulent zone (Marriott 1996). However, the amounts and grain sizes deposited depend on sediment supply and duration and depth of overbank flow.

The channel banks can gradually become the highest points on the floodplain as the thicker, coarser overbank deposits build up to form natural levees. The stream may then deposit sediment on its bed during high flows that do not exceed BANKFULL DISCHARGE, resulting in the normal river surface being above the level of the floodplain. Both natural levees and artificial embankments set away from the channel as macrochannel banks to a two-stage channel, afford some protection from flooding. In extreme cases, floodwater may break through the levee, forming a crevasse channel and washing sediment from the channel and reworked from the levee onto the floodplain. These sediments form a crevasse splay of coarser material than the underlying floodplain alluvium. Coarse material can also be transferred from the channel during overbank flow due to the action of convection currents set up by turbulence at bends in the channel. This is because the flow of water on the floodplain tends to travel directly down valley and at meanders the direction of flow within the channel is at an angle to that on the floodplain (Knight and Shiono 1996). Bedload sediment can then be picked up and spread in a lobe downstream from the outer bank of the bend.

In arid and semi-arid areas floodplains are formed mainly during major flood events (superfloods) with recurrence intervals in the region of 10,000 years. These floods bring in material from highland areas of the catchment. Between these events the sediment is reworked by the more frequent relatively minor flash floods that occur in the ephemeral channels rather than new material being added. Studies of the streams of central Australia (Pickup 1991) show different landforms depending on the scale of flooding that gave rise to them. The superfloods result in large sandsheets and sand threads and the contemporary macrochannel system which has levees, FLOODOUTS and floodbasins wherein sediment is reworked during flooding or by aeolian processes in dry periods.

Floodplain classification

Floodplains can be classified according to their morphology rather than the manner in which they were formed or the processes active at present. The floodplain as a whole, together with its parent stream, can be considered as a macroform, whereas the mesoforms are the component parts of the channel, e.g. bar forms, and the floodplain (levees, crevasse channels and splays, backswamps and oxbow lakes). Mesoforms can influence flow and deposition patterns on the macroform. In this classification the microforms are the small-scale structures superimposed on the mesoforms, for example, ripples, dunes, shrinkage cracks (Lewin 1978).

A further classification that takes into account the genesis of the floodplain was suggested by Nanson and Croke (1992) and is based on the relationship between the ability of streams to entrain and transport sediment and the resistance to erosion of the bank sediment. Three classes are recognized: high energy streams with non-cohesive banks, medium energy streams with non-cohesive banks and low energy streams with cohesive banks. Within these classes further levels of classification can be set up based on primary geomorphic factors such as channel cutting and filling and lateral accretion on point bars, and secondary factors such as scroll bar formation and organic (peat) accumulation. The primary geomorphic factors depend on stream power and sediment load and can therefore identify different environments for floodplain formation.

Floodplains respond to changes in channel processes that result from alterations in flow regime and/or sediment supply (Schumm 1977) though the response may be at a slower rate. The sedimentary record of these changes is often incomplete as it is often complicated by episodes of erosion, so although floodplain formation is polygenetic, much of the evidence is destroyed. It could be suggested that, over geological timescales, all floodplains are polygenetic as external influences such as climate and relative base-level change are not constant.

Floodplain sedimentation rates

In humid areas rivers flood every 1–2 years, though it is mainly extraordinary events that are studied and documented due to their catastrophic effects on human activity and property. Studies of flood frequency and sedimentation patterns have, therefore, been carried out with the aim of attempting to predict, and thus avoid, the destructive effects of extreme events. As explained above, the floodplains of many rivers are composed of sediments accumulated from channel and flood activity and many studies record the thickness of deposits in various parts of the floodplain by extraordinary events or they give the nature of the sediment deposited. Generally figures are highly variable and are only useful as a general guide to likely sedimentation rates in the various floodplain environments indicated.

Flooding is an essentially random occurrence and it is difficult to sample satisfactorily from the floodplain surface during a flood. However, borehole data, sediment cores, ^{14}C dating, pollen analysis and radioactive nuclides (e.g. ^{137}Cs and ^{210}Pb) can all be used to estimate sedimentation rates over a period of time. As an example, Brown (1987) used some of these techniques to estimate that sedimentation rates on the River Severn floodplain in the UK over the past 10,000 years have been around 1.4 mm per year.

Floodplains act as storage space or sediment sinks for alluvial sediment. While they are being stored the sediment may be reworked by fluvial, aeolian, biological and/or pedological agents, often over considerable periods of time. The stored sediment may subsequently be eroded and re-incorporated into the sediment budget of the drainage basin. The residence time of sediment in storage will vary according to factors such as surface topography, climate and vegetation and the relative return frequency of major flood events. Originally, sediment storage on floodplains was studied so that SEDIMENT BUDGETs of drainage basins could be calculated; now, however, interest in the storage of sediment that was originally part of the suspended load of the parent stream has increased with an awareness of the ability of contaminants such as heavy metal ions and radionuclides to adhere to and be transported with this fine material.

Summary

Understanding the processes involved in floodplain formation is important because of the interaction between human activity and floodplain environments. These processes include both within-channel and overbank processes which rely on the interaction between channel and floodplain flows during flooding, and which account for the distribution of different sediment grain sizes across the floodplain. Some recent work has investigated the rate of sedimentation and sediment storage on floodplains using a multiproxy approach. As floodplains act as sinks for alluvial sediments, this work is particularly useful for studies of contamination.

References

Brown, A.G. (1987) Holocene floodplain sedimentation and channel response of the lower River Severn, *Zeitschrift für Geomorphologie NF* 31, 293–310.

Knight, D.W. and Shiono, K. (1996) River channel and floodplain hydraulics, in M.A. Anderson, D.E Walling and P.D. Bates (eds) *Floodplain Processes*, 139–181, Chichester: Wiley.

Lewin, J. (1978) Floodplain Geomorphology, *Progress in Physical Geography* 2, 408–437.

——(1984) British meandering rivers; the human impact, in C.M. Elliott (ed.) *River Meandering*, 362–369, New York: American Society of Civil Engineers.

Markham, A.J. and Thorne, C.R. (1992) Geomorphology of gravel-bed river bends, in P. Billi, R.D. Hey, C.R. Thorne and P. Tacconi (eds) *Dynamics of Gravel Bed Rivers*, 433–450, Chichester: Wiley.

Marriott, S.B. (1996) Analysis and modelling of overbank deposits, in M.A. Anderson, D.E. Walling and P.D. Bates (eds) *Floodplain Processes*, 63–93, Chichester: Wiley.

Nanson, G.C. and Croke, J.C. (1992) A genetic classification of floodplains, *Geomorphology* 4, 459–486.

Nanson, G.C. and Young, R.W. (1981) Overbank deposition and floodplain formation on small coastal streams of New South Wales, *Zeitschrift für Geomorphologie NF* 25, 332–347.

Pickup, G. (1991) Event frequency and landscape stability on the floodplain systems of arid Central Australia, *Quaternary Science Reviews* 10, 463–473.

Schumm, S.A. (1977) *The Fluvial System*, New York: Wiley.

Wolman, M.G. and Leopold, L.B. (1957) River floodplains: some observations on their formation, *US Geological Survey Professional Paper* 282C, 87–107.

Further reading

Anderson, M.A., Walling, D.E. and Bates, P.D. (eds) (1996) *Floodplain Processes*, Chichester: Wiley.

Knighton, D. (1998) *Fluvial Forms and Processes: A New Perspective*, London: Edward Arnold.

Richards, K. (1982) *Rivers: Form and Process in Alluvial Channels*, London: Methuen.

SUSAN B. MARRIOTT

FLOW REGULATION SYSTEMS

Flow regulation systems is a general expression including all works constructed along a watercourse in order to regulate the flow in the channels. Flow regulation systems can be planned for maintaining steady-state conditions along a river or for avoiding uncontrolled erosion and/or sedimentation processes. On the other hand, they can be planned for obtaining a more constant distribution of the discharge in the channel by reducing both highwater and minimum flow peaks.

The devices utilized for avoiding accelerated erosion or deposition in the watercourse consist of hydraulic works capable of maintaining flow velocities within suitable ranges of values, which should be adequately chosen for each river. Besides this, a whole set of other measures are planned and constructed for opportunely directing the flow stream in order to avoid the occurrence of maximum velocities in proximity of the banks (see BANK EROSION). The most commonly used remedial measures consist in the direct protection of the banks by means of walls, sheet piles, prefabricated structures, gabions, anchored geotextiles, loose debris and biofixing plants (e.g. planting cuttings, sowing herbaceous species, etc.).

Groynes are typical works which direct the water flow; they are made up of structures stretching both downflow and with a certain angle with respect to the mean flow and can be either linear or composite (such as sledge hammers or bayonet joints). Furthermore, various arrangements of boulders in the river bed are works which increase friction by reducing velocity and increasing turbulence.

Finally, check dams are works which reduce excessive riverbed slope profiles. They break the river's profile into several stretches with lower inclinations, fixing the level of non-erodable weir crests and introducing artificial steps in the longitudinal profile of the watercourse. Besides being constructed with very heterogeneous materials, check dams can be of various patterns and dimensions. When they produce a real impoundment upstream they should rather be considered proper DAMs. To this regard, it should be mentioned that in the Alps many dams constructed for hydroelectric purposes work also as reservoirs for the regulation of water discharge during considerable highwater or minimum flow events. Apart from the regulation of flow, dams across great rivers usually have multipurpose functions, such as production of electric energy, navigation control, irrigation supply and flood control (Jansen *et al.* 1979).

Check dams are used also in watercourses capable of transporting and depositing large amounts of sediments (see BEDLOAD). In this case, they retain a certain amount of sediment in the upper part of the basins and reduce solid transport during highwater and minimum flow phases. Since the early 1950s various kinds of open weirs have been planned and constructed; among them, the most used ones are fissure-weirs and filtering weirs. The latter can be comb-like, with windows, network patterned, etc. (Figure 63). These particular check dams accomplish a two-fold purpose: (1) to retain most of the sediment during highwater phases and release it later during a low-flow phase; (2) to retain the coarse debris (including floating tree-trunks etc.) in order to avoid damage to the hydraulic structures downstream. Since the late 1980s check dams with low-angle filtering intake accompanied by a drainage gallery have been constructed along watercourses subject to overconcentrated flows ascribable to DEBRIS FLOWS.

In order to regulate the flow rate distribution during the year, other kinds of works are utilized. Among these, flood attenuation basins should be mentioned. They consist in hydraulic works which connect the watercourse to sufficiently large artificial basins capable of working as retaining reservoirs during river spates. These flood protection structures, also defined as flow-control weirs with energy dissipators, consist of a downstream regulation dam which allows the passage of a flow not superior to a prefixed value,

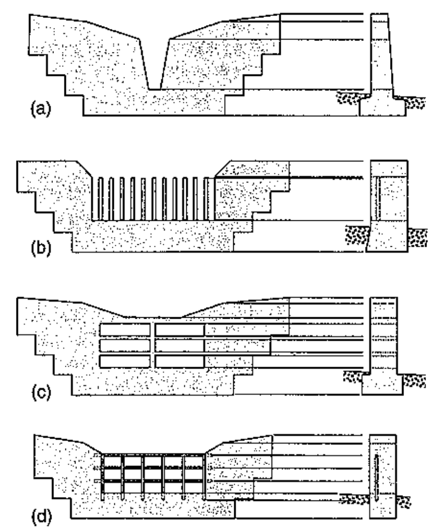


Figure 63 Examples of (a) fissure-weirs; and filtering weirs; (b) comb-like; (c) with windows; (d) network patterned

even in the occurrence of a greater flow. The exceeding volume of water is stored in the basin and returned to the river when the highwater phase is dwindling (Bell and Manson 1998). Flood protection structures can also be equipped with an upstream check dam in order to avoid discharge water from entering the flood attenuation basin when flow rate is below a certain average value. In this way the sedimentation of the bedload usually transported is avoided and, consequently, so is the progressive reduction of the basin's retention capacity. For example, along the central course of the Yangtze River, once this watercourse started to wander in its alluvial plain, a flood control weir was constructed which allows highwater flow to be diverted into a large dissipator basin provided along one of its tributaries (Jingjiang River and Dongting Lake). These hydraulic facilities, built in 1953 and extended in 1990, are made up of a 54-hole barrage, 1,054 m long, and can store up to 8,000 $m^3 s^{-1}$ of water in order to prevent inundations in the stretch located immediately downstream of the catchment works, thus protecting the town of Wuhan and the plain of Janghan from flooding.

Fillways and diversion canals are natural or, more often, artificial watercourses which receive a portion of a river's highwater in order to divert it into another basin or give it back to the same river, downstream of a critical stretch. Diversion canals are characterized by constant flow whereas fillways are utilized only occasionally. In the case of a diversion canal which subtracts water from a river and gives it back downstream of a critical point, usually an inhabited centre, the bypass canal should be long enough to avoid impoundment problems (raised hydrostatic levels due to a return effect starting from the confluence).

Other works which effect a reduction in flow rate levels, consist in modifications of the geometrical features of the river bed (see HYDRAULIC GEOMETRY). These solutions should be implemented with care as they are the result of engineering viewpoints and seldom do they take into sufficient account the geomorphological features of the canal. The usual procedure involves deepening, widening and redressing the canal. In the case of reshaping and widening of a canal's section, a decrease of the stream velocity takes place with consequent sedimentation and rising of the river bed which, therefore, requires periodical dredging. Other cases of course modification consist in canal straightening. For example, the best known cases of straightening are those in the lower course of the Mississippi River, which were carried out during the 1930s and 1940s by means of cutoffs of the meandering course for a total length of 210 km. The consequence of these modifications was an average increase of the riverbed gradient (Winkley 1982) which contributed to an increase of bedload, a rise of the canal's width/depth ratio and a tendency to change from a MEANDERING course to a braided (see BRAIDED RIVER) one. Thus, navigation problems arose due to the decreased depth of the river bed and, as a consequence, expensive dredging and bank protection works were necessary in order to mitigate the problem which still persists in this stretch of the river.

Among the works which regulate flow, artificial embankments should also be mentioned; they are the first fluvial works of a certain entity ever realized by humans. Indeed, it is well known that the construction of considerable embankments in the Po Plain, in northern Italy, started during the Roman age, although traces of partial embankments date back even to the previous Etruscan epoch (Marchetti 2002).

In the alluvial plains of economically advanced regions, flood protection works are connected to dense canal networks with draining and irrigation purposes which allow intensive farming activities in the plain areas and, at the same time, reduce the risk of flooding in urban areas.

Extreme cases of flow regulation systems consist in the implementation of real changes of the hydrographic network of a region by means of artificial diversions of important watercourses. In the ex-Soviet Union, in a vast territory characterized by north- and east-bound large rivers which flow through largely infertile cold regions and extensive drought-stricken areas, impressive diversions were planned and partially implemented. For example, the southward diversions of the rivers Peciora and Irtys, carried out in order to avoid the drying up of Lake Aral following the heavy water exploitation of Syrdarja and Amudarja, has produced, on the one hand, the drying up of vast northern territories covered by taiga and, on the other hand, the swamping of large areas to the south.

References

Bell, F. and Manson, T.R. (1998) The problems of flooding in Ladysmith, Natal, South Africa, in M. Eddleston and J.G. Manud (eds) *Geohazards and Engineering Geology*, Engineering Geology Special Publication 14.

Jansen, P., van Bendegom, L., Berg., J., de Vries, M. and Zanen, A. (1979) *Principles of River Engineering*, London: Pitman.

Marchetti, M. (2002) Environmental changes in the central Po plain (Northern Italy) due to fluvial modifications and man's activities, *Geomorphology* 44(3–4), 361–373.

Winkley, B.R. (1982) Response of the Lower Mississippi to river training and realignment, in R.D. Hey, J.C. Bathurst and C.R. Thorne (eds) *Gravel-bed Rivers*, 659–680, Chichester: Wiley.

Further reading

Brookes, A. (1985) River channellization: traditional engineering methods, physical consequences and alternative practices, *Progress in Physical Geography* 9, 44–73.

Church, M. (1995) Geomorphic response to river flow regulation: case studies and time-scales, *Regulated Rivers, Research and Management* 11, 3–22.

Gregory, K.J. (1995) Human activity and palaeohydrology, in K.J. Gregory, L. Starkel and V.R. Baker (eds) *Global Continental Palaeohydrology*, 151–172, Chichester: Wiley.

Petts, G.E. (1984) *Impounded Rivers: Perspectives for Ecological Management*, Chichester: Wiley.

Sear, D.A., Darby, S.E., Thorne, C.R. and Brookes, A.B. (1994) Geomorphological approach to stream stabilization and restoration: case study of the Mimmshall Brook, Hertfordshire, *Regulated Rivers, Research and Management* 9, 205–223.

SEE ALSO: flood; fluvial erosion quantification; river restoration

MAURO MARCHETTI

FLOW VISUALIZATION

The human eyes have some difficulty in perceiving the displacement of air, water and of most fluids. However, looking at fluid interfaces (surface boils in river) or particles suspended in a fluid (such as snow flakes or air bubbles) aids in the recognition of structured bodies within the moving fluid. Flow visualization provides a set of tools that have led to significant advances in our understanding of fluid dynamics. Flow visualization tools allow us to track and follow individual turbulent flow structures as they develop (Lagrangian reference frame) or to define flow parameters at one specific location within the flow, such as flow recirculation boundaries (Eulerian reference frame). As a result, these tools have become indispensable in the study of the structured motion of fluids on and within the globe's surface.

Sketches of turbulent flows made by Leonardo da Vinci (1452–1519) demonstrate that flow visualization has long fascinated the imagination of scientists. However, in order to complement quantitative flow monitoring, flow visualization techniques were developed primarily in laboratory studies of fluid mechanics during the last half-century. In the later 1960s, the use of flow visualization led to a major breakthrough in the understanding of turbulent boundary layer structure. Kline *et al.* (1967) (using air bubbles) and Corino and Brodkey (1969) (using neutrally buoyant particles) showed that the flow near a boundary, albeit turbulent in nature, exhibit structured patterns and mechanisms. Today, a wide range of visualization techniques is routinely used in laboratory studies (see the *Atlas of Flow Visualization* and the proceedings from several *International Symposiums on Flow Visualisation* for in-depth reviews of techniques and results from wind tunnels and water flumes experiments).

Most flow visualization techniques rely on the presence of a foreign tracer in the flow. These are often suspended particles or dye/smoke injected at specific locations. The use of tracers relies on the fact that the fluid motion can be inferred from the tracer movement matching that of the fluid. This implies a clear understanding of the mode of introduction or generation of the tracer in the flow, of the relationship between tracer and fluid motion, and of the physical significance of the observed tracer motion. In highly turbulent flows, cameras are used to aid the interpretation of flow patterns from moving particles. The use of dye/smoke present the advantage of providing a quick expression of the flow structure, but can diffuse rapidly away from the source of injection and, thus, reduce the tracing distance.

As well as in laboratory studies, flow visualization techniques can be used in aeolian, fluvial and costal environments where tracers often naturally occur. Matthes (1947), for example, provided an extensive classification of flow structures found in a river based on visual observations of surface boils, waves, turbidity differences and other visual indices. Roy *et al.* (1999) described two flow visualization techniques used in the natural environment. The first uses the turbidity difference to describe the development of flow structures at the shear layer between two merging streams. The other involves the injection of a milky white fluid to visualize shedding motions from the recirculating flow region in the lee of the obstacle.

Two different approaches of *quantitative flow visualization* exist. The first involves sampling velocity over a dense grid using single point velocity meters, such as electromagnetic current meters or acoustic Doppler velocimeters. This grid is then used to create maps of the turbulent parameters of interest (Bennett and Best 1995). As the measurements are not taken simultaneously, this approach provides a frozen picture of the general flow patterns. The second approach it to take velocity measurements using several single probes simultaneously. Such a set-up allows space–time velocity matrices to be created from which space–time velocity coherence and footprints of flow structures can be observed and described within the region covered by the velocity metres (Buffin-Bélanger *et al.* 2000).

The increasing use of multi-point velocity measurement techniques, such as acoustic Doppler profiling (Wewetzer *et al.* 1999) and particle image velocimetry (Bennett *et al.* 2002), is bound to create new breakthroughs in our understanding of flow structure. These techniques rely on the measurement of velocity from embedded particles in the moving fluid and allow the temporal variability of the flow to be described quantitatively at one point as well as the spatial variability of flow patterns in time. Hence, these techniques combine the qualitative realism of flow visualization with quantitative velocity measurements.

Our ability to use computational fluid dynamics (CFD) to simulate complex flows is increasing dramatically. This improvement gives rise to more and more sophisticated *numerical flow visualization* (Lane *et al.* 2002). Traditional flow visualization also complements CFD in allowing us to compare numerical results to natural flow behaviour.

References

Bennett, S.J. and Best, J.L. (1995) Mean flow and turbulence structure over fixed, two-dimensional dunes: implications for sediment transport and dune stability, *Sedimentology* 42, 491–514.

Bennett, S.J., Pirim, T. and Barkdoll, B.D. (2002) Using simulated emergent vegetation to alter stream flow direction within a straight experimental channel, *Geomorphology* 44, 115–126.

Buffin-Bélanger, T., Roy, A.G. and Kirkbride, A.D. (2000) On large-scale flow structures in a gravel-bed river, *Geomorphology* 32, 417–435.

Corino, E.R. and Brodkey, R.S. (1969) A visual investigation of the wall region in turbulent flow, *Journal of Fluid Mechanics* 37, 1–30.

Kline, S.J., Reynolds, W.C., Schraub, F.A. and Runstadler, P.W. (1967) The structure of turbulent boundary layers, *Journal of Fluid Mechanics* 30, 741–773.

Lane, S.N., Hardy, R.J., Elliot, L. and Ingham, D.B. (2002) High-resolution numerical modelling of three-dimensional flows over complex river bed topography, *Hydrological Processes* 16, 2,261–2,272.

Matthes, G.H. (1947) Macroturbulence in natural stream flow, *Transactions of American Geophysical Union* 28, 255–265.

Roy, A.G., Biron, P.M., Buffin-Bélanger, T. and Levasseur, M. (1999) Combined visual and quantitative techniques in the study of natural turbulent flows, *Water Resources Research* 35, 871–877.

Wewetzer, S.F.K., Duck, R.W. and Anderson, J.M. (1999) Acoustic Doppler current profiler measurements in coastal and estuarine environments: examples from the Tay Estuary, Scotland, *Geomorphology* 29, 21–30.

Further reading

Atlas of Flow Visualization, Vols 1, 2 and 3. Visualization Society of Japan (eds), Tokyo.

Van Dyke, M. (1988) *An Album of Fluid Motion*, Standford, CA: The Parabolic Press.

Tavoularis, S. (1986) Techniques for turbulence measurement, in *Encyclopedia of Fluid Mechanics*, 1,207–1,255, Texas: Gulf Publishing Company.

SEE ALSO: boundary layer

THOMAS BUFFIN-BÉLANGER AND ALISTAIR D. KIRKBRIDE

FLUIDIZATION

A geomaterial becoming a fluid, behaving like a fluid, usually associated with high-speed debris flows such as very mobile rockslides and NUÉE ARDENTE. It may be that fluidization within a rock avalanche mass can account for the properties of catastrophic landslides. The fluidization arises from the interaction of energetic particles and trapped air under pressure. In a fluidized system granular material is supported by air (or any other gas, but usually air). The geomorphological relevance of fluidization is to ground failure in which debris flows occur, usually as long run systems apparently supported by air. Some classic landslides, e.g. Blackhawk, Elm, Frank, Saidmarrah, etc., fall into this category. One of the most striking properties of debris is its relatively high fluidity; debris flows with 80–90 per cent granular solids by weight can move in sheets about 1 m thick over surfaces with slopes of 5–10 degrees. The high fluidity suggests that the debris is 'fluidized' in the sense that this term is used by the chemical engineers. In fluidization, the interstitial fluid moves so rapidly upwards through the granular solids that these solids are suspended.

There seem to be three possibilities for debris flow mechanisms: either the flow moves essentially as a mass, supported by an air cushion, which allows long travel; or the system is fluidized in the classical sense and air and particles are interacting to keep the system mobile; or the mobility is ensured by particles interacting with each other in the high energy system. There are factors which support all three views. The Blackhawk landslide fell about 1,000 m from the mountain of the same name in south California in some prehistoric period. Possibly the slide moved almost as a single unit, gaining a nearly frictionless ride on a cushion of compressed air. This theory appears to be supported by the marginal ridges of debris formed where material was dropped as air leaked from the edges of the slide mass, and by debris cones on the landslip formed by air leakages blasting up through holes in the main mass.

There were no witnesses to the Blackhawk event; the Elm rockslide in Switzerland was closely observed. In September 1881 a large part of the Plattenberg mountain fell about 400 m and landed near the village of Elm. A vast amount of rubble crashed to the valley floor, bounced 100 m up the opposite wall, turned and – in less than a minute – careered down the valley for over 1 km before coming to a sudden stop. It is feasible that the Elm rockslide could have travelled down the valley on a cushion of air but eyewitnesses suggest that some fluidization mechanism was more likely; the surface of the slide was observed 'boiling' in great turbulence, and parts of the slide ran into houses, suggesting a great overall mobility.

Further reading

Brunsden, D. and Prior, D.B. (eds) (1984) *Slope Instability*, Chichester: Wiley.

Shreve, R.L. (1968) *The Blackhawk Landslide*, Geological Society of America Special Paper 108.

Voight, B. (ed.) (1978) *Rockslides and Avalanches, 1: Natural Phenomena*, Amsterdam: Elsevier.

Waltham, T. (1978) *Catastrophe: The Violent Earth*, London: Macmillan.

SEE ALSO: liquefaction

IAN SMALLEY

FLUVIAL ARMOUR

'Armour' is one of several terms applied to clastic deposits in which the surface layer is coarser than the substrate (see BOULDER PAVEMENT). The phenomenon is widespread in gravel-bed rivers, where maximum surface and subsurface grain sizes may be similar but the respective median diameters differ by a factor of 2 to 4 because fine material is largely absent from the surface. It can also occur in sand-bed rivers that contain a little gravel, which becomes concentrated on the surface as a stable lag deposit. The traditional explanation is preferential winnowing of finer sediment from the surface during degradation, for example below dams which cut off the gravel flux from upstream so that the river erodes its bed to regain a capacity BEDLOAD. This degradation is self-limiting because as the surface coarsens, the transport capacity of the flow declines and the bed becomes immobile. This 'static' armouring has been investigated in flume experiments with no sediment feed and has been modelled mathematically; see Sutherland (1987) for a good review.

Coarse surface layers also exist in unregulated rivers with an ongoing sediment supply and peak flows which can transport all sizes of bed material. This 'mobile armour' allows the channel to be in equilibrium (neither degrading nor aggrading, neither coarsening nor fining) despite the size-selective nature of bedload transport, since intrinsically less mobile coarse fractions are preferentially available for transport whereas potentially mobile fine fractions are mainly hidden in the subsurface (Parker and Klingeman 1982). The armour forms by vertical winnowing during active bedload transport: entrainment of coarse clasts during floods creates gaps which are filled mainly by finer grains. Extreme floods may wash out the armour, but it re-forms during intermediate flows in most environments. In ephemeral streams there may be no such flows and armouring is generally absent (Laronne *et al.* 1994).

Mobile armour helps reduce bedload flux to match a restricted supply, with static armour as the limiting case when supply is cut off completely (Dietrich *et al.* 1989; Parker and Sutherland 1990). Changes in grain packing, as well as size distribution, are involved in this self-regulation; they include imbrication of coarser clasts and the development of pebble clusters, stone cells and transverse steps. The river bed is thus a degree of freedom in the adjustment of alluvial channels (see CHANNEL, ALLUVIAL) towards grade (see GRADE, CONCEPT OF).

References

Dietrich, W.E., Kirchner, J.F., Ikeda, H. and Iseya, F. (1989) Sediment supply and the development of the coarse surface layer in gravel-bedded rivers, *Nature* 340, 215–217.

Laronne, J.B., Reid, I., Frostick, L.C. and Yitshak, Y. (1994) The non-layering of gravel streambeds under ephemeral flood regimes, *Journal of Hydrology* 159, 353–363.

Parker, G. and Klingeman, P.C. (1982) On why gravel bed streams are paved, *Water Resources Research* 18, 1,409–1,423.

Parker, G. and Sutherland, A.J. (1990) Fluvial armor, *Journal of Hydraulics Research* 28, 529–544.

Sutherland, A.J. (1987) Static armour layers by selective erosion, in C.R. Thorne, J.C. Bathurst and R.D. Hey (eds) *Sediment Transport in Gravel-bed Rivers*, 141–169, Chichester: Wiley.

ROB FERGUSON

FLUVIAL EROSION QUANTIFICATION

The quantification of fluvial erosion, as for any other geomorphic process, is a way to make more precise and objective the assessment of the morphological changes that affect the Earth's relief. Beside this purpose, fluvial erosion quantification has a critical importance in the field of APPLIED GEOMORPHOLOGY because it can help the elaboration of erosion rate prediction models that are useful in the evaluation of GEOMORPHOLOGICAL HAZARDS.

Fluvial erosion consists of the entrainment and transport (by solution, suspension and traction) of particles that make up the stream bed or the stream banks (see EROSION). To quantify fluvial erosion, therefore, means to assess the amount of materials that a stream is capable of wearing away from bedrock or alluvial channels. These materials, however, become part of the total solid load and it is almost impossible to differentiate them from those delivered to the stream and deriving from the denudation processes acting within the whole basin. In other words, although streams are powerful erosional agents, they operate in conjunction with other exogenetic agents and with unconcentrated surface waters in particular; therefore drainage basins are acknowledged as the fundamental geomorphic units and rivers as their main elements along which energy is available. Consequently the quantification of fluvial erosion must be understood as quantification of the overall DENUDATION affecting both the slopes and the channels of the drainage basins.

The estimates of denudation affecting the slopes are often based on the direct determination of the amount of sediments removed from small sample areas or erosion field plots, or on the measurement of ground surface lowering and calculation of the volume of sediment dislodged. The evaluations of SOIL EROSION obtained in this way, however, are strictly dependent on the peculiarities of the studied slopes; therefore they have only local significance and can lead to misleading conclusions when they are extended to larger areas. In spite of this limit, a large number of field data on soil loss are useful in developing erosion prediction equations when they are plotted against several erosion controlling factors: the UNIVERSAL SOIL LOSS EQUATION is the most famous of them.

Channel erosion is usually evaluated by surveying the modifications of channel form and calculating the volume of material removed. The methods used include measurements to reference pegs and periodical controls of both channel cross section and long profile. Long-term variations can also be estimated by comparing aerial photographs of different periods.

As the material removed from slopes and channels of a drainage basin is the source for fluvial transport, the total amount of sediment load (see SEDIMENT LOAD AND YIELD) at the main river mouth can measure the intensity of denudation affecting the whole basin. Actually, most of the attempts directed towards the determination of erosion rate are based on the assessment of stream load quantity. Such assessment can be approached in different ways. One approach consists in the measurement of all the transported materials at the recording stations, that is to say the direct measurement of dissolved load, SUSPENDED LOAD and BEDLOAD. The indirect approaches, instead, lead to the stream load prediction through theoretical formulae or multiple regressions.

The field determination of dissolved load is obtained by portable instruments that measure certain water quality parameters, such as conductivity and pH. More often the dissolved solid content is determined in laboratories by evaporation of known volumes of water and weighing the residue. Suspended load concentration is obtained by measuring the turbidity of water samples collected by specially developed devices, ranging in complexity from simple dip-bottles to sophisticated apparatus; the total suspended load is then obtained multiplying the suspended load concentration by the discharge. The assessment of bedload is extremely difficult; many measuring apparatuses have been developed, like slot traps, collecting basin, basket samplers, etc., but none has been universally accepted as adequate for the determination of bedload.

Direct measurements of total solid load by rivers have encountered many problems; among them there are the high costs of instruments, the running expenses, and the alteration of pattern of flow and transport by the presence of the sampler, which can distort especially bedload data. One more problem is sampling both in time and space. Observations made at given time intervals could miss the extreme events; furthermore it may require many years of record before data are enough to be significant. The choice of sampling site must consider the accessibility of instruments, the lack of interference and the planning of a dense instrumentation network.

The theoretical estimation of solid load implies the derivation of specific formulae based on the characteristic of channel flow and of the transported materials. This procedure is unsuitable to predict suspended load, as it is essentially a non-capacity load, but it has been tentatively followed to predict bedload; however none of the derived formulae would seem to offer a completely satisfactory prediction.

An indirect method largely used to evaluate fluvial erosion takes into account the data available on suspended load (the most systematically measured at recording stations) and leads to significant regressions that relate suspended load to several parameters which express the principal factors influencing the spatial pattern of sediment production (Table 18). Once obtained these equations are used to predict suspended load of rivers lacking a recording station. Although suspended load values are a partial measure of erosion processes in drainage basins, they have been used also to obtain world maps of denudation rate (Fournier 1960).

Table 18 Some examples of multivariate regressions between suspended load and controlling variables

Author	Region	Equation
Fournier (1960)	Temperate alpine areas	$\log E = 2.65 \log (p^2 P^{-1}) + 0.46 H_m \tan \phi - 1.56$ E = suspended sediment yield (tonnes km^{-2} $year^{-1}$); p = precipitation in month of maximum precipitation (mm); P = mean annual precipitation (mm); H_m = mean elevation of basin (m); ϕ = mean basin slope (°)
Jansen and Painter (1974)	Humid microthermal climatic areas	$\log S = -5.073 + 0.514 \log H + 2.195 \log P - 3.706 \log V + 1.449 \log G$ S = suspended sediment yield (tonnes km^{-2} $year^{-1}$); H = altitude (m.a.s.l.); P = mean annual precipitation (mm); V = measure of vegetation cover; G = estimate of proneness to erosion
Jansen and Painter (1974)	Temperate climates	$\log S = 12.133 - 0.340 \log Q + 1.590 \log H + 3.704 \log P + 0.936 \log T - 3.495 \log C$ S = suspended sediment yield (tonnes km^{-2} $year^{-1}$); Q = annual discharge (10^3 m^3 km^{-2}); H = altitude (m a.s.l.); P = mean annual precipitation (mm); T = average annual temperature (°C); C = natural vegetation index
Ciccacci *et al.* (1986)	Italy	$\log TU = 2.79687 \log D + 0.13985\ \Delta a + 1.05954$ $r^2 = 0.96128$ Tu = suspended sediment yield (tonnes km^{-2} $year^{-1}$); D = drainage density (km km^{-2}); Δa = hierarchical anomaly index

References

Ciccacci, S., Fredi, P., Lupia-Palmieri, E. and Pugliese, F. (1986) Indirect evaluation of erosion entity in drainage basins through geomorphic, climatic and hydrological parameters, *International Geomorphology*, 2, 33–48, Chichester: Wiley.

Fournier, F. (1960) *Climat et érosion: la relation entre l'érosion du sol par l'eau et les précipitations atmosphériques*, Paris: Presses Univ. de France.

Jansen, J.M.L. and Painter R.B. (1974) Predicting sediment yield from climate and topography, *Journal of Hydrology* 21, 371–380.

Further reading

Cooke, R.U. and Doornkamp, J.C. (1990) *Geomorphology in Environmental Management*, Oxford: Clarendon Press.

Gregory, K.J. and Walling D.E. (1973) *Drainage Basin Form and Process*, London: Edward Arnold.

ELVIDIO LUPIA-PALMIERI

FLUVIAL GEOMORPHOLOGY

Fluvial geomorphology is strictly the geomorphology of rivers. As rivers have always held a prominent role in the study of landforms, it is not surprising that debates about fluvialism, as to whether rivers could produce their valleys, continued to rage early in the nineteenth century until uniformitarianism prevailed, whence temperate areas were seen as the result of rain and rivers. At the end of the nineteenth century rivers were seen as central to the Davisian normal cycle of erosion which came to exert a dominant influence upon geomorphology for the first half of the twentieth century (Gregory 2000). It took time to appreciate that there was insufficient understanding of fluvial processes and, with hindsight, it has been suggested that the way in which G.K.Gilbert approached rivers, including significant contributions on the transport of debris by flowing water (Gilbert 1914), could have provided at least an additional, if not an alternative, approach to that developed by Davis.

Until the mid-twentieth century fluvial geomorphology was dominated by attempts to interpret landscapes in terms of phases of river evolution with emphasis placed upon terraces as indicating sequences of valley development, and erosion or planation surfaces employed to reconstruct stages of landscape development. Questioning the basis for such reconstructions, and realizing a need for a greater focus upon fluvial processes, was answered by *Fluvial Processes in Geomorphology* (Leopold *et al.* 1964) which came to have a dramatic influence upon the way in which fluvial geomorphology was subsequently pursued. In addition to increasing interest in hydrological processes (see HYDROLOGICAL GEOMORPHOLOGY) and leading to recognition that the drainage basin could be regarded as the fundamental geomorphic unit (Chorley 1969), it provided the basis for an expansion of research on the contemporary fluvial system. Emphasis was effectively placed upon seven different themes (Gregory 1976) which were: drainage network morphometry; drainage basin characteristics particularly in relation to statistical models of water and sediment yield; links between morphology and process in the hydraulic geometry of river channels; and the controls upon river channel patterns; together with theoretical approaches to the fluvial system; investigations of the significance of dynamic contributing areas in runoff generation; and finally, ways of analysing changes in fluvial systems including the PALAEOHYDROLOGY-river metamorphosis approach. At this stage in the rapid development of fluvial geomorphology, textbooks were produced including emphases on dynamics and morphology (Morisawa 1968), form and process (Richards 1982; Morisawa 1985; Knighton (1984), 1998), rivers and landscape (Petts and Foster 1985), and the fluvial system (Schumm 1977a). There was thus a progressive development of fluvial geomorphology; in his conclusion to *The Fluvial System*, Schumm (1977a) contended that landscape, like science itself, proceeds by episodic development. In the second part of the twentieth century the development of fluvial geomorphology proceeded episodically, and the result of research progress made has been to demonstrate, in turn, exactly how episodic fluvial systems can be.

Comparatively few explicit definitions of fluvial geomorphology have been given in papers and books but five (Table 19) indicate core themes and to some extent intimate how the subject has evolved, expanded and progressed since the 1960s.

In the course of the development of fluvial geomorphology since the middle of the twentieth century, three sequential phases can be visualized: one of import and expansion, one of consolidation, and one of innovation. The first saw import and utilization of understanding and techniques from other disciplines including hydraulics, hydrology, sedimentary geology and engineering. Clarification and refinement of field approaches together with modelling methods including stochastic, deterministic (Werritty 1997) and experimental (Schumm *et al.* 1985) approaches enabled the developing foundation to focus upon equilibrium concepts throughout several distinct sub-branches of fluvial geomorphology established as the outcome of this phase. The independence or dependence of variables involved in research investigations was anticipated to vary according to the steady, graded, or cyclic timescale (Schumm and Lichty 1965) of the investigation being considered. Once established with a significant number of practitioners, fluvial geomorphology experienced a second phase, one of internal consolidation, characterized by investigations of changes over time as referred to different timescales, embracing palaeohydrology, and river channel adjustments as instigated by the effects of land use changes impacting upon the channel. These investigations meant that controls upon change of the fluvial system were explored, including thresholds, complex response and sensitivity. The third phase of innovation, aided by new technology of remote sensing and GPS, has seen the development of exciting links between investigations undertaken at several spatial scales, together with equally innovative developments linking studies of process with landscape development. This phase has been one of export whereby results from fluvial geomorphology are contributing to multidisciplinary projects and making a distinctive input to management problems (e.g. Thorne *et al.* 1997).

Against this background of development of the subject, a choice for fluvial geomorphology was suggested to exist by Smith (1993) because he perceived the discipline to be at a crossroads, requiring major changes in ways of thinking and operating, so that he proposed it needed to move forward and to adopt the ways of the more competitive sectors of the Earth and biosciences. However this need may have been overstated (Rhoads 1994) in view of the vitality shown by recent publications in fluvial geomorphology, and by the way in which collaboration and multidisciplinary activity have increased, complemented by attention being devoted to the scientific foundation of the subject. Rhoads (1994: 588) sees the most critical challenge facing fluvial geomorphologists as that of devising effective strategies for integrating a diverse assortment of research, spanning a broad range of spatial and temporal scales.

From the ideas that prompted Smith's challenge and Rhoads's response, it is possible to seek a general definition; the broad range of research

Table 19 Some definitions of fluvial geomorphology

Definition	Source
Fluvial geomorphology has as its object of study not only individual channels but also the entire drainage system	Kruska and Lamarra (1973) cited by Schumm (1977b)
A primary objective of fluvial geomorphology must be to contribute to *explanation* of relationships among the physical properties of flow in mobile-bed channels, the mechanics of sediment transport driven by the flow, and the alluvial channel forms created by spatially differentiated sediment transport	Richards (1987)
Geomorphology is the study of Earth surface forms and processes; fluvial phenomena – those related to running water	Graf (1988)
The study of changing river channels is the domain of fluvial geomorphology. ... Fluvial geomorphology is a field science; classification and description are at the heart of this science	Petts (1995)
The science that seeks to investigate the complexity of behaviour of river channels at a range of scales from cross sections to catchments; it also seeks to investigate the range of processes and responses over a very long timescale but usually within the most recent climatic cycle	Newson and Sear (1998), cited by Dollar (2000)

approaches that exists; and ways in which collaboration is now possible. Embracing earlier proposals (Table 19) a general definition for fluvial geomorphology could be that it investigates the fluvial system at a range of spatial scales from the basin to specific within channel locations; at timescales ranging from processes during a single flow event to long-term Quaternary change; undertaking studies which involve explanation of the relations among physical flow properties, sediment transport and channel forms; of the changes that occur both within and between rivers; and that it can provide results which contribute in the sustainable solution of river channel management problems.

Although developed at different times and progressed significantly since the seven themes suggested in 1976 (Gregory 1976), there is now a range of research approaches which are the branches of fluvial geomorphology occupying most practitioners at any one time. These can now be envisaged as focused on components of the fluvial systems, process mechanics, temporal change and management applications.

Components of the fluvial system

Studies of components of the fluvial system have been concerned particularly with morphology of elements of that system across the range of spatial scales from in-channel locations to the complete drainage basin. Particularly significant investigations focused on relations between form and process in fluvial systems and upon the controls upon morphology. This has required definitions which can be applied in different basins including those for channel capacity, channel planform, floodplain extent, and drainage density of the channel network; and at each of these levels there have been attempts to establish equilibrium relations between indices of process and measurements of fluvial system form. Some of the earliest developments in fluvial geomorphology were concerned with analysis of *drainage networks* using techniques of drainage basin morphometry and with the relationship between *channel capacity* and the frequency of the bankfull discharge which was thought to exercise a major control upon channel morphology. In these and other components of the fluvial system it has now been appreciated that the links between form and process and the associated explanation is more complex than at first thought. Thus drainage networks could not easily be related to discharge and channel processes, and the relationship between channel capacities and controlling discharge has been the subject of considerable research, particularly the way in which networks generate the Geomorphic unit hydrograph (Rodriguez-Iturbe and Rinaldo 1998). Relations between dimensions of cross-sectional area and width can be used to provide a basis for discharge estimation at ungauged sites (Wharton *et al.* 1989). In addition research has focused upon *river channel patterns*, upon the controls on single thread and multi-thread patterns and what determines the thresholds between them. The floodplain is also controlled by the interaction between recent hydrological and sediment history together with the characteristics of the local area; and the variability of floodplain characteristics has been reflected in the definition of the river corridor as well as of the floodplain itself. Three major floodplain classes, based on stream power and sediment characteristics, have been recognized (Nanson and Croke 1992), further subdivided into a combination of thirteen floodplain orders and suborders, namely:

1 High energy non-cohesive floodplains: disequilibrium landforms which erode either completely or partially as a result of infrequent extreme events.
2 Medium energy non-cohesive floodplains: in dynamic equilibrium with the annual to decadal flow regime of the channel and not usually affected by extreme events. Preferred mechanism of floodplain construction is by lateral point bar accretion or braid channel accretion.
3 Low energy cohesive floodplains: usually associated with laterally stable single-thread or anastomosing channels. Formed primarily by vertical accretion of fine-grained deposits and by infrequent channel avulsion.

As more is known about each of the several spatial scales of investigation of the fluvial system it is appreciated that the question of explanation relies upon the controls that apply to each particular spatial level. Thus it is necessary to see how the flows and sediment transport are significant in relation to each of the spatial levels of the fluvial system, and how they interrelate. Furthermore, when focusing on the integrity of the fluvial system, it is necessary to consider how a hierarchy of interrelated components makes up the river basin channel structure. Any

such structure needs to take account of the progress made by biologists and aquatic ecologists in this regard and an original framework (Downs and Gregory 2003: Chapter 3) involves seven nested scales which are drainage *basin*, basin *zones*, valley *segments*, stream *reaches*, *channel unit*, *within channel*, and channel *environment at a point*. In addition there are a number of environmental flows that have been defined (Dollar 2000) including those that maintain a channel morphology and which have been specified for the purposes of practical application.

Process mechanics

Process mechanics began with hydrological analyses whereby fluvial processes were examined from the standpoint of analysis of stream hydrographs and their generation, so that investigations of dynamic contributing areas became a major reason for field experiments based in small experimental catchments. On the basis of the considerable progress made, attention then moved to the sediment budget, and to suspended sediment, bedload and solute loads. Analysis previously dominated by simple rating curves, which assumed a linear relationship between suspended sediment concentration and discharge, was refined once it was shown that because of hysteretic effects and sediment supply problems, the relationships were much more complex so that earlier estimates of rates of denudation had to be revised. Bedload transport had been very difficult to measure so that transport equations often tended to assume a capacity load, whereas along many rivers the transport of material was supply limited. Advances in instrumentation enabled continuous recording devices to be used providing the basis for more complete explanatory analysis of transport rates; and continuous measurements of channel bank erosion could be the basis for more precise relationships between erosion rates and the controlling variables. Such studies facilitated more detailed investigations within the channel and these have been concerned with the entrainment of bed material, the patterns of flow and sediment movement at confluences, and the controls upon in-channel change. A particularly fruitful theme has derived from the interrelationships with ecology because aquatic ecologists have investigated river channels in relation to instream habitat conditions and aquatic plant distributions; combination of such results with geomorphological data has promoted biogeomorphological investigations of river channels. Along rivers bordered by riparian trees, or flowing through forested areas, the investigation of CWD (coarse woody debris) has become important because such CWD exerts an influential control upon the channel processes, the morphology and ecological characteristics. Numerous investigations have been undertaken considering the impact of CWD upon channel morphology, demonstrating the extent and significance of wood often as debris dams, together with the impact on channel processes, the reasons for spatial variations, as well as the stability of dams and their persistence together with the management implications.

Temporal change

Study of temporal changes had been a long tradition in fluvial geomorphology through the interpretation of past stages of development based on river terrace sequences, but it was not until 1954 that the idea of palaeohydrology was proposed (Leopold and Miller 1954). This contrasted with earlier approaches because it was retrodictive in approach, utilizing understanding that had been gained of contemporary processes, and it was exemplified by Quaternary palaeohydrology suggested by Schumm (1965) and augmented by ideas of river metamorphosis (Schumm 1977a). Palaeohydrology evolved (Gregory 1983) to utilize knowledge of contemporary processes applied to the past, whereas river metamorphosis similarly employed contemporary relationships between channel form and process as a basis for interpreting river channel resulting from a range of causes including dam and reservoir construction, land use change including urbanization and, particularly, as a consequence of channelization. Such human-induced channel changes were found to be extensive (e.g. Brookes and Gregory 1988) and were superimposed upon the impact of shifts in sequences of climate which in some parts of the world, such as Australia, led to the alternation of periods of drought-dominated and flood-dominated regimes. Analysis of river channel changes was initially confined to particular reaches affected, often downstream from the influencing factor, but they have subsequently been analysed in the context of the entire basin with attention accorded to the spatial distribution of adjusting channels and emphasis given to the extent to which they are potentially able to recover to their former condition (e.g. Fryirs and Brierley 2000). The considerable progress achieved as a result of studies of channel change includes ways in which thresholds can be identified, reaction and relaxation times (Graf 1977), patterns of palaeohydrological change in different parts of the world (Benito and Gregory 2003) all culminating in further, more informed understanding of the ways in which fluvial processes relate to environmental change (Brown and Quine 1999) and how fluvial systems and sedimentary sequences reflect shifts of climate, both short or longer term during the Quaternary (Maddy *et al.* 2001). A particularly effective way in which studies of the past have been successful has been the analysis of palaeofloods (see HYDROLOGICAL GEOMORPHOLOGY), overcoming not only inaccuracies in estimating the ages of floods and in reconstructing flood discharges, but also allowing incorporation of palaeoflood data into flood frequency analysis, in order to analyse the effects of climatic shifts and non-stationarity. The results of palaeoflood hydrology have been of practical application by bringing very specific benefits in the design or retrofitting of dams or other floodplain structures.

Management applications

Palaeoflood analysis is just one example of ways in which fluvial geomorphology provides applications to management. Many other applications have been developed and initially were very problem and reach specific, including estimation of sediment yield and the possibility of gullying and channel change (Schumm 1977a), progressing through consequences of particular impacts such as channelization (Brookes 1988), leading to improved procedures for management (Brookes and Shields 1996) and then to comprehensive statements of ways in which fluvial geomorphology can be applied to river engineering and management (Thorne *et al.* 1997). Particular emphasis has been placed upon river restoration and how fluvial geomorphology is able to contribute significantly to restoration projects (Brookes 1995). One aspect of restoration to be considered is what is natural (Graf 1996) and therefore what should be the objective for a particular restoration project. It is being appreciated that in all cases where human impact affects fluvial systems, current knowledge of the way in which the system has evolved can illuminate the way in which management is undertaken. It is also the case that some aspects of human activity are now being substantially reversed and the implications of dam removal (Heinz III Center 2002) is one topic of current interest.

It may seem from the foregoing outline of approaches in fluvial geomorpology that they have become increasingly reductionist and diverse, but there are a number of ways in which there has been integration and collaborative activity tending to unify fluvial geomorphology – although remembering that fluvial geomorphology is not as independent of other disciplines as it once was. Thus analysis of sediment slugs showing how waves of sediment are transmitted through a fluvial system has implications for management, and for interpretation of temporal change as well as for the understanding of contemporary channels' forms and processes. In addition, the use of fallout radionuclides including Cs-137 and Pb-210 has enabled precise dating of specific fluvial changes including floodplain sedimentation (e.g. Walling and He 1999). Based upon knowledge of a number of specific cases, the limits of explanation and prediction have been emphasized, including ten ways to be wrong (Schumm 1991). This introduces the idea of risk so that the fluvial system can be seen in terms of the incidence of twenty-eight geomorphic hazards which may occur, associated with drainage networks, hillslopes, main channels, piedmont and plain areas (Schumm 1988). From each of the above themes, clear signs are emerging of more integrated investigations, for example linking process and morphology in bank stability and modes of channel adjustment which involve bank erosion. There is great awareness of, and interaction between, the range of spatial scales investigated. Such linking analysis is now being facilitated by enhanced remote sensing techniques, and GPS which enhances the detail of data capture and the speed of analysis. Progress is now being made towards enhanced conceptual models which seek to model aspects of the fluvial system in ways not previously possible (Coulthard *et al.* 1999). The outstanding challenge is for further understanding to be achieved of the way in which information can be linked from one timescale to another.

It is inevitable that the investigation of rivers should have expanded greatly, as one of the most researched fields of geomorphology; it is now becoming more integrated but not necessarily strictly confined to geomorphology, as links with ecology, engineering and hydraulics and sedimentary geology prove to be very worthwhile, and multidisciplinary approaches are increasingly common. The concerns expressed by Smith (1993) are being met, and fluvial geomorphology is now sufficiently well founded to address major questions including what has been suggested to be the greatest challenge: 'to understand the way in which short timescale and small space scale processes operate to result in long timescale and large space scale behaviour' (Lane and Richards 1997: 258).

References

Benito, G. and Gregory, K.J. (2003) *Palaeohydrology: Understanding Global Change*, Chichester: Wiley.
Brookes, A. (1988) *Channelized Rivers. Perspectives for Environmental Management*, Chichester: Wiley.
——(1995) River channel restoration: theory and practice, in A. Gurnell and G. Petts (eds) *Changing River Channels*, 369–388, Chichester: Wiley.
Brookes, A. and Gregory, K.J. (1988) Channelization, river engineering and geomorphology, in J.M. Hooke (ed.) *Geomorphology in Environmental Planning*, 145–168, Chichester: Wiley.
Brookes, A. and Shields, F.D. (eds) (1996) *River Channel Restoration. Guiding Principles for Sustainable Projects*, Chichester: Wiley.
Brown, A.G. and Quine, T. (eds) (1999) *Fluvial Processes and Environmental Change*, Chichester: Wiley.
Chorley, R.J. (1969) The drainage basin as the fundamental geomorphic unit, in R.J. Chorley (ed.) *Water, Earth and Man*, 77–100, London: Methuen.
Coulthard, T.J., Kirkby, M.J. and Macklin, M.G. (1999) Modelling the impacts of Holocene environmental change in an upland river catchment, using a cellular automaton approach, in A.G. Brown and T. Quine (eds) *Fluvial Processes and Environmental Change*, 31–46, Chichester: Wiley.
Dollar, E.J. (2000) Fluvial Geomorphology, *Progress in Physical Geography* 24, 385–406.
Downs, P.W. and Gregory, K.J. (2003) *River Channel Management*, London: Arnold.
Fryirs, K. and Brierley, G. (2000) A geomorphic approach to the identification of river recovery potential, *Physical Geography* 21, 244–277.
Gilbert, G.K. (1914) The transportation of debris by running water, *US Geological Survey Professional Paper* 86.
Graf, W.L. (1977) The rate law in fluvial geomorphology, *American Journal of Science* 277, 178–191.
——(1988) *Fluvial Processes in Dryland Rivers*, Berlin: Springer-Verlag.
——(1996) Geomorphology and policy for restoration of impounded American rivers: what is 'natural'? in B.L. Rhoads and C.E. Thorn (eds) *The Scientific Nature of Geomorphology*, 443–473, Chichester: Wiley.
Gregory, K.J. (1976) Changing drainage basins, *Geographical Journal* 142, 237–247.
——(1983) (ed.) *Background to Palaeohydrology*, Chichester: Wiley.
——(2000) *The Changing Nature of Physical Geography*, London: Arnold.
Heinz III Center (2002) *Dam Removal. Science and Decision Making*, Washington: The Heinz Center.
Knighton, D. (1998) *Fluvial Forms and Processes*, 2nd edition, London: Arnold.
Kruska, J. and Lamarra, V.A. (1973) Use of drainage patterns and densities to evaluate large scale land areas for resource management, *Journal of Environmental Systems* 3, 85–100.
Lane, S.N. and Richards, K.S. (1997) Linking river channel form and process: time, space and causality revisited, *Earth Surface Processes and Landforms* 22, 249–260.
Leopold, L.B. and Miller, J.P. (1954) Postglacial chronology for alluvial valleys in Wyoming, *United States Geological Survey Water Supply Paper* 1,261, 61–85.
Leopold, L.B., Wolman, M.G. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: Freeman.
Maddy, D., Macklin, M.G. and Woodward, J.C. (eds) (2001) *River Basin Sediment Systems: Fluvial Archives of Environmental Change*, Amsterdam: Balkema.
Morisawa, M.E. (1968) *Streams: Their Dynamics and Morphology*, New York: McGraw-Hill.
——(1985) *Rivers: Form and Process*, London: Longman.
Nanson, G.C. and Croke, J.C. (1992) A genetic classification of floodplains, *Geomorphology* 4, 459–486.
Newson, M.G. and Sear, D.A. (1998) The role of geomorphology in monitoring and managing river sediment systems, *Journal of the Chartered Institution of Water and Environmental Management* 12, 18–24.
Petts, G.E. (1995) Changing river channels: the geographical tradition, in A. Gurnell and G. Petts (eds) *Changing River Channels*, 1–23, Chichester: Wiley.
Petts, G.E. and Foster, I.D. (1985) *Rivers and Landscape*, London: Arnold.
Rhoads, B.W. (1994) Fluvial geomorphology, *Progress in Physical Geography* 18, 588–608.
Richards, K.S. (1982) *Rivers: Form and Process in Alluvial Channels*, London: Methuen.
——(1987) Fluvial geomorphology, *Progress in Physical Geography* 11 432–457.
Rodriguez-Iturbe, I. and Rinaldo, A. (1998) *Fractal River Basins*, Cambridge: Cambridge University Press.
Schumm, S.A. (1965) Quaternary Palaeohydrology, in H.E. Wright and D.G. Frey (eds) *The Quaternary of the United States*, 783–794, Princeton: Princeton University Press.
——(1977a) *The Fluvial System*, New York: Wiley.
——(1977b) Applied fluvial geomorphology, in J.R. Hails (ed.) *Applied Geomorphology*, 119–156, Amsterdam: Elsevier.
——(1988) Geomorphic hazards – problems of prediction, *Zeitschrift für Geomorphologie Supplementband* 67, 17–24.

Schumm, S.A. (1991) *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge: Cambridge University Press.
Schumm, S.A. and Lichty, R.W. (1965) Time, space and causality in geomorphology, *American Journal of Science* 263, 110–119.
Schumm, S.A., Mosley, M.P. and Weaver, W.E. (1985) *Experimental Fluvial Geomorphology*, Chichester: Wiley.
Smith, D.G. (1993) Fluvial geomorphology: where do we go from here? *Geomorphology* 7, 251–262.
Thorne, C.R., Hey, R.D. and Newson, M.D. (eds) (1997) *Applied Fluvial Geomorphology for River Engineering and Management*, Chichester: Wiley.
Walling, D.E. and He, Q. (1999) Changing rates of overbank sedimentation on the floodplains of British rivers during the past 100 years, in A.G. Brown and T. Quine (eds) *Fluvial Processes and Environmental Change*, 207–222, Chichester: Wiley.
Werritty, A. (1997) Chance and necessity in geomorphology, in D.R. Stoddart (ed.) *Process and Form in Geomorphology*, 312–327, London: Routledge.
Wharton, G., Arnell, N.W., Gregory, K.J. and Gurnell, A.M. (1989) River discharge estimated from channel dimensions, *Journal of Hydrology* 106, 365–376.

KENNETH J. GREGORY

FOLD

Structures that originally were planar, like a sedimentary bed, but which have been bent by horizontal or vertical forces in the Earth's crust. Folds can occur at scales that range from mountain ranges to small crumples only a few centimetres long. They can be gentle or severe, depending on the nature and magnitude of the applied forces and the ability of the beds to be deformed. When beds are upfolded or arched (Figure 64a) they are called *anticlines*, whereas downfolds or troughs are termed *synclines*. A *monocline* is a step-like bend in otherwise horizontal or gently dipping beds. Folds can be asymmetrical in shape and, when the deformation is particularly intense, they can be overturned. They also tend to occur in groups rather than in isolation and some mountains consist of a folded belt. Folds can result from various processes: compression in the crust, uplift of a block beneath a cover of sedimentary rock so that the cover becomes draped over the rising block, and from gravitational sliding and folding where layered rocks slide down the flanks of a rising block and crumple. The process of folding may involve either small-scale shearing along many small fractures or flowage by plastic deformation of the rock.

Large folds can have a substantial influence upon landform development. Folding can also occur rapidly, creating vertical increases in elevation of up to 10 m 1,000 y^{-1}. However, when folds cease to grow, the influence of erosion becomes increasingly more important than the original shape of the fold. Thus, in the case of an anticline, initially it will form an area of upstanding, often donal relief. However, as it is eroded, the uppermost strata may be cut through. If the older rock that forms the core is of low resistance the result is a *breached anticline* in which a series of inward-facing escarpments rise above a central lowland. A classic example of this is the Weald of southern England (Jones 1999) (Figure 64b). Conversely, as in the Paris and London Basins, the rim of a syncline will possess outward facing scarps. Folding has major impacts on river systems. If a fold develops across a stream course the river, if it can cut down quickly enough, can maintain its course transverse to the developing structure. Such a channel is said to be *antecedent*.

Antecedence is, however, only one explanation for drainage that cuts across anticlinal structures. An alternative model (Alvarez 1999) is that fold ridges emerge from the sea in sequence, with the erosional debris from each ridge piling up against the next incipient ridge to emerge, gradually extending the coastal plain seaward. The new coastal plain, adjacent to each incipient anticline, provides a level surface on which a newly elongated river could cross the fold, positioning it to cut a gorge as the fold grew. This mechanism is in effect a combination of antecedence and superimposition. The model has been applied both to the Appalachians of the USA and the Apennines of Italy.

On actively developing anticlinal folds, drainage density varies according to the gradient of the evolving slopes. However, the form of the relationship between gradient and drainage density is process-dependent. Talling and Sowter (1999) suggest that a positive correlation occurs when erosion results from overland flow, while a negative correlation occurs when erosion is dominated by shallow mass-wasting. A traditional description of drainage in folded areas such as the Jura is provided by Tricart (1974).

Where sedimentary rocks are tilted by folding there may be a succession of lithologies exposed that have differing degrees of resistance. River channel incision will tend to be more effective on

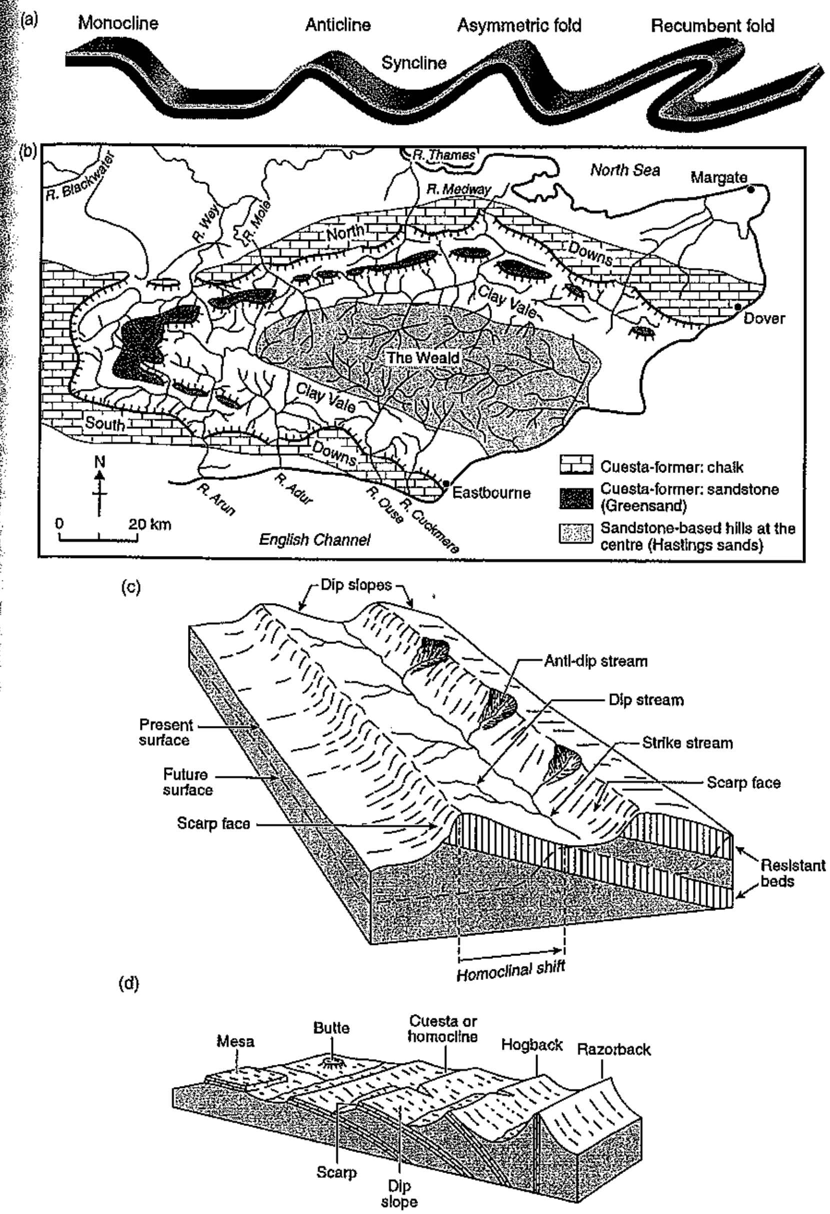


Figure 64 Folds and their relationship to relief: (a) some major types of fold; (b) the Wealden Anticline of south-east England; (c) drainage and slope forms associated with dipping strata; (d) drainage and slope forms associated with strata of progressively steepening dip

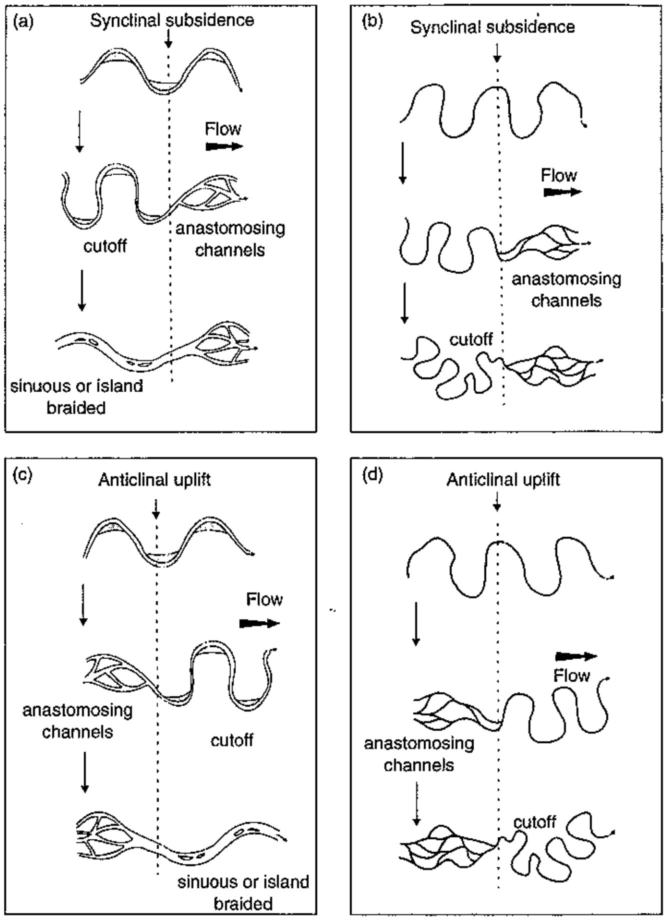


Figure 65 River channel patterns associated with synclinal and anticlinal activity: (a) and (c) = mixed-load meandering channels; (b) and (d) = suspended-load meandering channels (modified from Ouchi in Schumn *et al.* 2000, figures 3.10, 3.11, 3.12 and 3.13)

less resistant beds leading to the development of a *strike valley* (Figure 64c), flanked on the up-dip side by a dip slope and on the down-dip side by an escarpment. The roughly parallel strike streams will be joined at high angles by short *dip streams* and *anti-dip streams*. As downward incision occurs, the rivers will migrate laterally by a process known as *homoclinal shifting*.

The angle of dip also influences topographic form (Figure 64d). Resistant beds in very gently dipping or horizontal beds, form flat-topped plateaux (MESAS). Modestly dipping beds create a CUESTA, whereas steeply dipping strata produce a ridge known as a HOGBACK.

In recent years tectonic geomorphologists (see, for example, Burbank and Anderson 2001) have

taken a great interest in how the style and rate of folding affects landform evolution. In particular, Schumm *et al.* (2000) have described how terrace formation, channel form, the locations of degradation and aggradation, valley long profiles, and the spatial distribution of flooding, may be related to folding activity. Figure 65 shows the response of stream channel form to anticlinal uplift and synclinal subsidence for mixed-load and suspended-load streams. Streams flowing across zones of uplift (live anticlines) may show deformed terraces and convex sections of long profile.

References

Alvarez, W. (1999) Drainage on evolving fold-thrust belts: a study of transverse canyons in the Apennines, *Basin Research* 11, 267–284.
Burbank, D.W. and Anderson, R.S. (2001) *Tectonic Geomorphology*, Oxford: Blackwell Science.
Jones, D.K.C. (1999) On the uplift and denudation of the Weald, *Geological Society of London Special Publication* 162, 25–43.
Schumm, S.A., Dumon, J.F. and Holbrook, J.M. (2000) *Active Tectonics and Alluvial Rivers*, Cambridge: Cambridge University Press.
Talling, P.J. and Sowter, M.J. (1999) Drainage density on progressively tilted surfaces with different gradients, Wheeler Ridge, California, *Earth Surface Processes and Landforms* 24, 809–824.
Tricart, J. (1974) *Structural Geomorphology*, London: Longman.

A.S. GOUDIE

FORCE AND RESISTANCE CONCEPT

Materials at the Earth's surface are subjected to a whole range of applied forces (inputs of energy) both physical and chemical. Whereas in Newtonian mechanics applied force (or stress) and reaction (or strain) are thought of as equal and opposite and as essentially simultaneous, the outcome of an application of unit force to Earth materials often cannot be readily predicted. The asymmetry between energy input and response which is almost universally experienced in geomorphology is due partly to the multifaceted nature of resistance and partly to the significance of the specific sequence of energy inputs.

B.W. Sparks was especially concerned to stress that the resistance of rocks to geomorphic processes was always contingent upon the precise manner in which energy was applied. An obvious example is the very different resistance of limestones under chemical or mechanical attack (cf. Sparks 1971). This difficulty is well exemplified by M.J. Selby's elaborate quantification of the ROCK MASS STRENGTH of hillslopes, which nevertheless can only be invoked when the applied force environment is closely defined (Selby 1993: 104). Similarly, the difficulty in writing rational physical equations to describe fluvial flow and sediment transport may be directly traced to the non-uniform expression of channel boundary resistances.

Coupled with this first asymmetry is the widespread observation that energy inputs of equal magnitude do not result in equal amounts of geomorphic work. The non-linearity (see NON-LINEAR DYNAMICS) of process and response is of profound significance in all historical sciences. The most common manifestation of this asymmetry in geomorphology is in hysteresis loops of discharge and sediment plots. The applied force of a given discharge will generally neither entrain nor carry the same volume or calibre of sediment on the rising and falling limbs of a flood. The asymmetry is produced by the variable quantity and quality of sediment available to be moved: that is, on the temporally specific state of the channel boundary. The study of slope failures also (cf. Schumm and Chorley 1964) provides examples of small inputs of energy which propel a system across a resistance threshold (see THRESHOLD, GEOMORPHIC).

References

Schumm, S.A. and Chorley, R.J. (1964) The fall of Threatening Rock, *American Journal of Science* 262, 1,041–1,054.
Selby, M.J. (1993) *Hillslope Materials and Processes*, 2nd edition, Oxford: Oxford University Press.
Sparks, B.W. (1971) *Rocks and Relief*, London: Longman.

SEE ALSO: Goldich weathering series; magnitude–frequency concept; non-linear dynamics; rock mass strength; threshold, geomorphic

BARBARA A. KENNEDY

FOREST GEOMORPHOLOGY

Forest geomorphology is a specialized area of research focused both on the interaction between forest ecosystem dynamics and landform processes, and the effects of forest management activities on the rates and thresholds of geomorphic process events. As a specialty within the field of geomorphology, the study of forest associated process dynamics has broad application to the study of ecosystem dynamics, endangered species and conservation studies, and paleo-landscapes.

Several features distinguish geomorphic processes in forest environments from those in other vegetation types. Forest ecosystems play a prominent role in many aspects of sediment production, transport and storage. Entrainment of soil by windthrow of trees, the binding effect on soils and regolith by tree root strength, and soil displacement by burrowing animals, are among many examples of how sediment movement is influenced by forests. Standing trees and fallen logs on slopes in forested terrain act as temporary sediment storage 'traps', while forest management activities, such as road construction and the removal of trees, transport and disturb sediment production in other ways. Forest composition plays a major role in the interception, evapotranspiration, infiltration, and runoff of precipitated water, and, as in the sediment examples cited, profoundly influences the type, frequency and mechanisms controlling mass wasting and slope stability. Ultimately, such influences of forest vegetation profoundly affect the development and operation of drainage basins, adding a significant organic component to sediment load, hydrologically influencing the discharge response of the channel network, even influencing channel development through bank resistance or flow deflection. Added to these considerations are natural cycles within the forest ecosystem such as wildfire, floods, disease or insect impacts, and human disturbance factors, which can affect the linkage of geomorphic processes, as well as their magnitude and frequency.

The diversity of objectives, approaches and professional disciplines of those who conduct forest-related process studies has resulted in poor communication among scientists and land managers who share many common interests. In recognition of the need to promote interdisciplinary dialogue, and approach complex environmental research within a spatial and temporal scale appropriate to the changes in forest succession affecting process-response, the International Council of Scientific Unions created the International Geosphere–Biosphere Program (IGBP) which recognizes several prominent forest biomes. Geomorphologists are placed prominently among the scientific teams that continue to illuminate the linkages between landforms and ecosystems.

History and development of forest geomorphology

While late nineteenth and early twentieth-century landscape theory focused on denudation and landform development, early geomorphologists were aware of variation in landscape appearance under varying climate and vegetation conditions. Cotton (1942) explains variations in landscape form as climatic 'accidents', including the resultant vegetation influences, as complications of the 'Geographical Cycle' theorized by William Morris Davis (1899). In a similar vein, Birot (1968) further expands upon Davis, and adds vegetation and soil factors to illustrate the influences of climatic variation. Peltier (1950) also refers to Davis's cycle, but places emphasis on the variations in geomorphic process activity under a variety of forest biomes (selva, savanna, boreal). Peltier credits Professor Kirk Bryan with many of these concepts of 'vegetation modified process'.

Hack and Goodlett (1960) illustrated the relationships between process and forest structure to identify relict features within a forested landscape, and heralded the concept of coexisting ecological and geomorphic equilibriums contributing to landform genesis in humid temperate forests, while Douglas (1968) described the effects in humid tropical forests, and added human disturbance factors. Chorley used examples such as these to illustrate the benefits of multidisciplinary approaches in geomorphology using a 'systems' approach to integrate information from widely divergent sources at different temporal and spatial scales. Chorley's 1962 work encouraged scientists to make contributions across disciplinary lines, and US Geological Survey ecologist Sigafoos (Sigafoos and Hendrick, 1961; Sigafoos 1964; Hupp and Sigafoos 1982) dated trees to determine the temporal and spatial activity of glaciers, floods and blockfields (Alestalo 1971). Successive works were absorbed by the resource management community, which funded research concentrated on the effects of timber harvest practices, road and bridge construction and land use change.

Hydrology, sediment budgets and channel stability in forested watersheds

The application of concepts and techniques from geomorphology to ecosystem studies and terrain analysis represents a great opportunity for the discipline, given the need for an interdisciplinary approach to complex environmental problems.

Large-scale forest management, such as practised by government agencies in the United States, has resulted in numerous controversies between the economic, recreation and conservation interests. In the northwestern Unites States, the practice of large-scale total tree removal, called 'clearcutting', is controversial because of ecological, hydrological and erosion concerns. The US Department of Agriculture, accustomed to years of soil erosion studies at its experiment stations, created a network of experimental forests in the 1960s. Paired watershed studies were conducted to evaluate resulting sediment and water yields resulting from various management treatments. Fredriksen (1970) described the effects of traditional 'clearcutting' using roads, tree removal using a unique cable system to completely suspend the trees as they were cut, and a control basin that remained fully forested. This brief research report sparked considerable interest both in the USA and elsewhere, and a number of research investigations have followed the recovery progress of these small watersheds in the H.J. Andrews experimental forest in western Oregon over the past fifty years (Swanson and Jones 2001).

The natural and management effects of large woody debris on channel morphology and sediment transport in forested streams has been of considerable research interest. Natural log debris in stream channels often produces 'log steps' or 'organic knickpoints' that produce a 'step-pool' profile in mountainous forest streams. These pools act as sediment and nutrient traps, provide fish and invertebrate habitats, and may persist for a long time (Swanson and Lienkaemper 1978). The study of channel stability in forested streams has taken on special significance due to the effects that disturbance within such channels has upon the spawning cycle of anadromous fish. Several species of salmon, steelhead and char have been listed as 'endangered' because of loss of spawning gravel habitat, or loss of access to headwaters. The origin of gravel spawning beds, and their preservation, has occupied substantial geomorphic research including the delivery of gravel from regolith by mass wasting, winnowing of fines by floods, the effects of woody debris on in-channel storage, and the effects of surface 'armour layer' development on entrainment and transport. Stream ecologists, working with fluvial geomorphologists, can provide insights into physical habitat processes and sensitivity to disturbance. Several outstanding classification and inventory schemes have been produced regionally to predict habitat sensitivity and stability (Brussock *et al.* 1985).

Forested slopes have been shown to exhibit lower runoff, increased interflow, and greater stability arising from the root strength of the trees. In general, soils are subsequently deeper than on unforested slopes under similar conditions, resulting in conditions that 'trigger' mass wasting events of both natural and man-induced origins (O'Loughlin 1974). Moss and Rosenfeld (1978) have shown that mass wasting events have the potential to alter the composition and character of the forest community structure in predictable ways, thus leading to a model of interrelationships between landform features and vegetation community characteristics.

On a larger scale, Caine and his co-workers (Swanson *et al.* 1988) demonstrated that ecosystem behaviour can be predicted by a better understanding of how landforms affect those processes. They illustrate that ecosystem–terrain interactions often take multiple forms, with patterns imposed by one set of interactions often coexisting in time and space with other sets. The linkage between ecosystem development and landform stability incorporates both geomorphic and biological events of varying magnitude and frequencies, such as wildfire, floods and landslides. Rosenfeld (1998) illustrates that threshold events, such as exceptional storms, can have predictable 'triggering' effects based on morphology, forest composition and management history. Thus, anticipating the effects of global change on forest biomes are realistic objectives.

Recognition of the complex interdisciplinary nature of landform–forest interactions, and the significance of these linkages in the assessment of human impacts and global change, has been included in the principal themes of the Earth Systems Science Committee, established by the National (US) Aeronautics and Space Administration in 1986. These themes have been incorporated in the international Earth Observation System, and in global research designs for the International Space Station. Geomorphologists will continue to be integral members of ground-based research teams quantifying the linkages between terrestrial processes and forest ecosystems. The US National Academy of Sciences has established 'Long Term Ecological Reserves', with a minimum research planning term of two hundred years. Other nations have expressed similar plans, and a global network of sites, focused on major forest biomes, is a major scientific objective.

As forest geomorphology becomes established as a significant sub-field within the discipline, the need for interdisciplinary education has become apparent. Several sessions dedicated to forest geomorphology have been held by the International Association of Geomorphologists (IAG), and at least one formal graduate programme has been established.

References

Alestalo, J. (1971) Dendrochronological interpretation of geomorphic processes, *Fennia* 105, 1–140.

Birot, P. (1968) *The Cycle of Erosion in Different Climates* (English trans.), London: B.T. Batsford.

Brussock, P.P., Brown, A., and Dixon, J.C. (1985) Channel form and ecosystem models, *Water Resources Bulletin* 21, 859–866.

Chorley, R.J. (1962) Geomorphology and general systems theory, *US Geological Survey Professional Paper* 500B, 1–10.

Cotton, C.A. (1942) *Climatic Accidents in Landscape Making*, 2nd edition, New York: Wiley.

Davis, W. M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.

Douglas, I. (1968) Natural and man made erosion in the humid tropics of Australia, Malaysia, and Singapore, Publication of Staff Members, Centre for SE Asian Studies, University of Hull, 2nd Series, No. 2, 17–29.

Fredricksen, R.L. (1970) Erosion and sedimentation following road construction and timber harvest on unstable soils in three small western Oregon watersheds, *USDA Forest Service Research Paper PNW-104*.

Hack, J. and Goodlett, J.C. (1960) Geomorphology and forest ecology of a mountain region in the Central Applachians, *US Geological Survey Professional Paper* 347.

Hupp, C.R. and Sigafoos, R.S. (1982) Plant growth and block-field movement in Virginia, *US Forest Service, General Technical Report* PNW-141, 78–85.

Moss, M.R. and Rosenfeld, C.L. (1978) Morphology, mass wasting and forest ecology of a post-glacial re-entrant valley in the Niagara escarpment, *Geografiska Annaler* 60A, 161–174.

O'Loughlin, C.L. (1974) A study of tree root strength deterioration following clearfelling, *Canadian Journal of Forest Research* 4, 107–114.

Peltier, L.C. (1950) The geographical cycle in periglacial regions as it is related to climatic geography, *Annals of the Association of American Geographers* 40, 219–236.

Rosenfeld, C.L. (1998) Storm induced mass-wasting in the Oregon Coast Range, USA, in J. Kalvoda, and C. Rosenfeld (eds) *Geomorphological Hazards in High Mountain Areas*, 167–176, Dordrecht: Kluwer.

Sigafoos, R.S. (1964) Botanical evidence of floods and flood-plain deposition, *US Geological Survey Professional Paper* 485-A.

Sigafoos, R.S. and Hendrick, E.L. (1961) Botanical evidence of the modern history of Nisqually Glacier, Washington, *US Geological Survey Professional Paper* 387-A.

Swanson, F.J. and Lienkaemper, G.W. (1978) Physical consequences of large organic debris in Pacific Northwest streams, *USDA Forest Service General Technical Report* PNW-69.

Swanson, F.J. and Jones, J.A. (2001) Geomorphology and Hydrology of the H.J. Andrews Experimental Forest, Blue River, Oregon. PNW Forest Range Experimental Station, Portland. Authors describe numerous studies focused on sediment yield and routing, water yield, nutrient flux and re-vegetation studies conducted in these experimental watersheds from 1953 to present. This information is updated at the following internet address: http://www.fsl.orst.edu/lter

Swanson, F.J., Krantz, T.K., Caine, N. and Woodhouse, R.G. (1988) Landform effects on ecosystem patterns and processes, *Bioscience* 38, 92–98.

Further reading

Deithich, W.E., Dunne, T., Humphrey, N.F. and Reid, L.M. (1982) Construction of sediment budgets for drainage basins, *USDA Forest Service General Technical Report* PNW-141, 5–24.

Froehlich, H.A. (1973) Natural and man-caused slash in headwater streams, *Logger's Handbook*, vol. 33, Portland: Pacific Logging Congress.

Harden, D., Ugolini, F. and Janda, R. (1982) Weathering and soil profile development as tools in sediment budget and routing studies, *USDA Forest Service General Technical Report* PNW-141, 150–154.

Kelsey, H.M. (1982) Hillslope evolution and sediment movement in a forested headwater basin, *USDA Forest Service General Technical Report* PNW-141, 86–96.

Marston, R.A. (1982) The geomorphic significance of log steps in forest streams, *Annals of the Association of American Geographers* 72, 99–108.

SEE ALSO: applied geomorphology; bar, river; channel, alluvial; climatic geomorphology; landslide; mass movement; sediment routing; step-pool system; threshold, geomorphic

CHARLES L. ROSENFELD

FORMATIVE EVENT

An important idea in geomorphological science is that morphological stability or stasis can be interrupted by brief, instantaneous, episodes of erosion or deposition when significant morphological change takes place (Erhart 1955; Butler 1959; Ager 1976; Gould 1982; Reading 1982; Dott 1983).

A second important idea is that all geomorphological processes are made up of discrete events of varying frequency, magnitude, duration and sequencing characteristics. If we are to understand landform change it can only be with respect to the characteristics of the events which cause change (Brunsden 1996).

An event is a period of activity of a process, at any place. Events may be classified according to their role in landform evolution. An effective event is one that exceeds the resistance or tolerance of a system and does work. Following Wolman and Gerson (1978) this is measured by the ratio of the event to the mean annual condition of erosion, denudation rate or deposition. Small but frequent events cause morphological change in a cumulative way. All that is required is time. A crucial component is the sequence in which events of different potential effectiveness occur. A very effective event may have considerable feedback effect on succeeding events. If all the available work has been done, later events may perform below their energy potential. If the effective event unlocks potential energy (e.g. by creating steep slopes) it may build in to the system further progressive and diffusive change. If the event prepares a threshold (see THRESHOLD, GEOMORPHIC) condition and is followed by another effective event there may follow unusual or rare forms of change.

It is therefore helpful to use the term formative event. A formative event is an event, of a certain frequency and magnitude, which controls the form of the land. If it does more work than the cumulative everyday event the landform it produces will persist (perhaps for long periods) despite the modifying effects of the more frequent events. It may require another formative event to obliterate the landforms produced or such an event may reinforce the effect. Multiple glaciation of a valley is an example, the 'U' form, once produced by a glacial 'event', may remain for millions of years, surviving all changes in the environmental controls. The word 'persistence' describes the length of time a landform survives as a diagnostic element of the landform assemblage.

References

Ager, D.V. (1976) *The Nature of the Stratigraphic Record*, London: Macmillan.

Brunsden, D. (1996) Geomorphological events and landform change. The centenary lecture to the Department of Geography, University of Heidelberg, *Zeitshrift für Geomorphologie NF* 40, 273–288.

Butler, B.E. (1959) Periodic phenomena in landscapes as a basis for soil studies, Melbourne, CSIRO, *Soil Publication* 14.

Dott, R.H. Jr. (1983) Episodic sedimentation: how normal is average? How rare is rare? Does it matter? *Journal of Sedimentary Petrology* 53 (1), 5–23.

Erhart, H. (1955) 'Biostasie' et 'rhexistasie' ésquisse d'une théorie sur le rôle de la pédogenèse en tant que phénomène géologique, *Comptes Rendus Academie de Sciences* 241, 1,218–1,220.

Gould, S.J. (1982) Darwinism and the expansion of evolutionary theory, *Science* 216, 385–387.

Reading, H.G. (1982) Sedimentary basins and global tectonics, *Proceedings Geologists' Association* 93, 321–350.

Wolman, M.G. and Gerson, R. (1978) Relative scales of time and effectiveness in watershed geomorphology, *Earth Surface Processes and Landforms* 3, 189–208.

DENYS BRUNSDEN

FRACTAL

Sciences such as geomorphology are concerned with inherently variable phenomena – things that are not exactly predictable or repeatable because of the sensitive dependence on initial conditions that many geomorphological systems exhibit (i.e. CHAOS THEORY and the 'butterfly effect'), and because of the contingencies among the elements that make up the system (see SELF-ORGANIZED CRITICALITY). While detailed long-term predictions of such systems are impossible, we should be able to provide some statistical bounds within which the future (or past) should lie. Moreover, while actual events may be unpredictable, they are obviously not unexplainable. Fractals are fundamental components of the methods that are required when analysing or modelling such systems, methods that are amenable to complex, non-linear dynamical systems (see COMPLEXITY IN GEOMORPHOLOGY).

Consider that our primary evidence of the past lies in the patterns we observe today – be it on hillslopes or in river valleys. If the means by which we attempt to characterize those patterns are not able to capture the true complexities of the systems, then how are we to turn back the hands of time and develop an understanding of the Earth's past (Werner 1995)? Fractal patterns, chaos and self-organization can provide the null hypotheses against which process-based interpretations can be tested.

The field of fractals emerged primarily from the writings of one person – Benoit Mandelbrot (1967, 1982) – to become a mainstream field of research. Along with chaos theory and self-organization, it has dramatically altered our view of nature, and of geomorphology (Turcotte 1992).

Fractals are the unique patterns left behind by the unpredictable movements of the world at work. The branching patterns of rivers and trees, the coastline of Britain, the pebbles in a stream, the spatial distribution of earthquake epicentres – all these can exhibit fractal patterns. Fractal objects show similar details on many different scales. Imagine, for example, the rough bark of a tree viewed through successively more powerful magnifications. Each magnification reveals more details of the bark's rugosity. In many geomorphological phenomena, such as river networks and coastlines, this fractal self-similarity has long been observed (e.g. Burrough 1981). This means that, as we peer deeper into a fractal image, the shapes seen at one scale are similar to the shapes seen in the detail at another scale. Fractals are formally defined as objects that are self-similar (Baas 2002).

The measure which most people use to quantify fractal scaling and self-similarity is the fractal dimension or D. The fractal dimension is a number that reflects the way in which the phenomenon fills the surrounding space. The fractal dimension of an object is a measure of its degree of irregularity considered at (theoretically) all scales, and it can be a fractional amount greater than the classical geometrical dimension of the object. The fractal dimension is related to how fast the estimated measurement of the object increases as the measurement device becomes smaller. A higher fractal dimension means the object is more irregular, and the estimated measurement increases more rapidly. For objects of classical geometry, such as lines or curves, the geometric or topological dimension and the fractal dimension are the same. The quantification of fractal patterns led to the discovery that many phenomena, when plotted using appropriate transformations, can be described using a power law ($1/f$ systems).

An important concept tied to the fractal dimension is that of spatial autocorrelation. If nearby conditions on a surface are very similar to each other, then we call that positive spatial autocorrelation. If nearby conditions on the surface are the opposite, then we call that negative spatial autocorrelation. Spatial autocorrelation is zero when there is no apparent relation between nearby conditions. A low fractal dimension for a surface (e.g. 2.1) indicates that the self-similarity exhibits high positive spatial autocorrelation. A high fractal dimension (e.g. 2.9) indicates that the self-similarity exhibits high negative spatial autocorrelation. A fractal dimension in the middle of the range (2.5) indicates that no spatial autocorrelation exists. Brownian motion is a classic example of a fractal at the middle range – it is a process with zero memory of where it came from and no knowledge of where it will go next.

Although, theoretically, labelling something a fractal implies that it exhibits self-similarity across all scales, in fact most natural objects possess a limited form of self-similarity – between certain limits or resolutions, the object behaves in a fractal-like manner. These are often called fractal elements, and it is possible that an object may possess multiple fractal elements. Many scientists now consider the boundaries at which fractal behaviour is observed to be important, for those boundaries clearly distinguish process limits. However, does knowledge about the limits to the form of a phenomena necessarily allow us to make statements about the limits of the process which is responsible for creating that form? The answer to that question remains unanswered, although it is at the heart of most fractal research.

One of the main reasons for the increased interest in the fractal dimension (D) is the awareness that dissipative dynamical systems and fractal spaces (and time) are linked – that we now have a theoretical basis with which to link form (e.g. D) and process (e.g. self-organized criticality). The lack of such a link has long been one of the criticisms levelled at fractal studies (e.g. Mark and Aronson 1984), so the discovery that a link can be made is an important step forward in fractal research. However, while self-organized critical models developed in a computer have been very successful at mimicking many varied systems, the unequivocal existence of self-organized criticality in real systems has yet to be confirmed. Geomorphic concepts, such as negative feedback, static equilibrium, and the concept of the graded stream, are all similar to the concept of self-organized criticality. These existing concepts provide an explanation for many geomorphic phenomena without the need to invoke a mechanism such as self-organized criticality. Many geomorphometric measures also are not statistically related to the fractal dimension – that alone indicates that fractals are not capturing all the aspects of a landscape that geomorphologists consider important (Klinkenberg 1992).

Earthquakes and avalanches are two of the more visible manifestations of self-organized critical systems. Their statistical properties, such as size distributions, generally obey power-scaling laws – they follow a fractal distribution (Bak 1996). If a form is found to be a fractal form then certain statistical properties follow. A fractal form has no one scale dominant – it is scale invariant – and its second moment is theoretically infinite. Conversely, a form which is not scale-invariant, a 'Gaussian' form, can be completely described by a few statistical moments. A fractal form will be characterized by rare intermittent events that, from a process point of view, dominate the statistical record. Thus, one of the challenges that fractal studies are attempting to meet is the characterization of such statistically intractable events or forms (e.g. Xu *et al.* 1993). Furthermore, such statistical properties mean that obtaining enough data with which to compare the predictions of models against reality is a not an easy process (Baas 2002).

Fractal concepts have been applied extensively in fluvial geomorphology (e.g. Rodriguez-Iturbe and Rinaldo 1997); there are several different aspects which can be studied. The most obvious is: what is the true length of a river? One could also, while considering the entire basin, examine the form of the river network within the basin. At a higher resolution, the actual planform of the river can be considered (i.e. quantifying sinuosity). At these scales one must not only consider the river itself, but also the river valley form and its effects on the geometry of the river. Going even further down the scaling hierarchy, studies of the fractal characteristics of river bedforms can also be made.

It has been found that many allometric relations observed in nature are not dimensionally consistent (Church and Mark 1980). For example, dimensional analysis would conclude that the length of the mainstream channel of a river should be proportional to the square root of the area of the basin. Most studies have observed that, in fact, the mainstream channel length is proportional to the 0.6th root of the area. Mandelbrot interpreted that as a fractal finding: if a river meanders such that it has a fractal dimension of 1.2, then the length–area relation (known in the literature as Hack's relation) should be to the 0.6th power (1.2 divided by 2).

Power laws, which are the signature of fractals, have been experimentally observed over a wide range of scales in probability distributions describing river basin morphology. Some of the observed fractal distributions have been:

- Horton's power laws of bifurcation and length.
- Stream lengths follow a power-law distribution.
- The cumulative total drainage area contributing to any link follows a power-law distribution.
- The mean of the local slope of the links of a drainage network scales in a fractal manner as a function of the cumulative area.

The fact that 'fractal' rivers exist in so many regions implies that fractal growth processes occur in every environment.

If we accept that river networks and topography can sometimes be characterized as fractals, then we must question why that occurs and what processes are responsible. The simplistic explanation is that scale-invariant form is the result of scale-invariant processes (e.g. Burrough 1981). Does this necessarily mean that a scale-invariant process operates over all scales, as the scale-invariant spatial form appears similar over all scales? We know that this can't be the case – consider the processes such as chemical weathering, frost action and soil creep that operate only at the microscale level. Obviously, the assumption of a one-to-one correspondence between the scale of the form and the scale of the process cannot hold. Self-organization provides the means of getting around this assumption. Large- and small-scale spatial structures emerge through the operation of small-scale processes. Simple rules at one level can lead to complex behaviour at a higher level, behaviour which is referred to as emergent behaviour. We do not have to program in the complex behaviour; it just appears as a consequence of the actions of the agents at the smaller scale.

Fractals provide an out from the constraints of Euclidean geometry, and capture the patterns of nature in an intuitive way. Experiments have shown how our perceptions of roughness agree very well with the measured fractal dimension of the object. Fractal geometry has shifted research agendas: while strict quantitative measurement, measurement that values quantifiable features like distance and degrees of angles, is still important, it is now recognized that measures also need to embrace the qualities of things – their texture complexity and holistic patterning. Chaos, self-organization and fractals have allowed us to step away from simple linear deterministic models and step towards models which capture the essence of predictably unpredictable natural systems.

References

Baas, A.C.W. (2002) Chaos, fractals and self-organization in coastal geomorphology: simulating dune landscapes in vegetated environments, *Geomorphology* 48, 309–328.

Bak, P. (1996) *How Nature Works*, New York: Copernicus.

Burrough, P.A. (1981) Fractal dimensions of landscapes and other environmental data, *Nature* 294, 240–242.

Church, M. and Mark, D.M. (1980) On size and scale in geomorphology, *Progress in Physical Geography* 4, 342–390.

Klinkenberg, B. (1992) Fractals and morphometric measures: is there a relationship? *Geomorphology* 5, 5–20.

Mandelbrot, B.B. (1967) How long is the coast of Britain? Statistical self-similarity and fractal dimensions, *Science* 156, 636–638.

——(1982) *The Fractal Geometry of Nature*, San Francisco: Freeman.

Mark, D.M. and Aronson, P.B. (1984) Scale-dependent fractal dimensions of topographic surfaces: an empirical investigation, with applications in geomorphology, *Mathematical Geology* 16, 671–683.

Rodriguez-Iturbe, I. and Rinaldo, A. (1997) *Fractal River Basins (Chance and Self-Organization)*, Cambridge: Cambridge University Press.

Turcotte, D.L. (1992) *Fractals and Chaos in Geology and Geophysics*, Cambridge: Cambridge University Press.

Werner, B.T. (1995) Eolian dunes: computer simulation and attractor interpretation, *Geology* 23, 1,107–1,110.

Xu, T., Moore, I.D. and Gallant, J.C. (1993) Fractals, fractal dimensions and landscapes – a review, *Geomorphology* 8, 245–262.

BRIAN KLINKENBERG

FRAGIPAN

A natural subsurface horizon found deep in the soil profile, which has been altered by pedogenic processes responsible for restricting the entry of water and roots into the soil matrix. Fragipans possess a higher bulk density than the horizons above, contain very little organic matter, are brittle when moist and exhibit slaking properties when immersed in water. Thickness ranges from 15–200 cm, enough to allow sufficient impact upon plant growth so that roots and water are unable to penetrate 60 per cent of the horizon. Fragipans develop mostly in mid-latitude, medium texture, acid materials overlying albic or argillic soil horizons, and with udic or aquic moisture regimes. Fragipans occur mainly beneath forest vegetation, in cultivated or virgin soils within various parent materials including glacial drift, loess, colluvium, lacustrine deposits and alluvium, though they are not found in calcareous deposits. Fragipans consistently possess an abrupt upper boundary at a depth of 30–100 cm beneath the ground surface (Witty and Knox 1989), and often exhibit evidence of soil formation.

The origins of fragipans are poorly understood, though three main formation mechanisms exist. These are: physical ripening during desiccation of initially slurried material; clay bridging; and bonding by an amorphous component (including Si, Al, and Fe). Unfortunately, fragipan is a generally poorly defined term, with many examples of fragipans worldwide unidentified due to their vague definition in the field.

Reference

Witty, J.E. and Knox, E.G. (1989) Identification, role in soil taxonomy, and worldwide distribution of fragipans, in N.E. Smeck and E.J. Ciolkosz (eds) *Fragipans: Their Occurrence, Classification, and Genesis*, SSSA Special Publication Number 24, 1–10.

Further reading

Smeck, N. E. and Ciolkosz, E. J. (eds) (1989) *Fragipans: Their Occurrence, Classification, and Genesis*, SSSA Special Publication Number 24.

STEVE WARD

FREEZE–THAW CYCLE

A freeze–thaw cycle is a cycle in which temperature fluctuates both above and below 0 °C. Field measurements indicate that most freeze–thaw cycles per year occur in the climates of low annual temperature range, which are dominated by diurnal or cyclonic fluctuations (French 1996: 26). These conditions are met in subpolar oceanic locations (e.g. Jan Mayen in the northern hemisphere, Kerguelen Islands or South Georgia in the southern hemisphere) and in intertropical high mountain environments (e.g. Andes, East Africa mountains). Among all cold environments, the least number of freezing and thawing days occur at high latitude and in continental climates, which

are dominated by seasonal temperature regimes. In all areas, most cycles occur in the upper 0–5 cm of the ground and only the annual cycle occurs at depths in excess of 20 cm.

Freeze–thaw cycles have important effects on soils, like FROST HEAVE, frost sorting or frost creep. It is important to distinguish between seasonally frozen ground and PERMAFROST (Washburn 1973: 15). In non-permafrost environments, the depth of seasonal freezing increases with increasing latitude, the range being from a few millimetres to more or less 3 metres. In permafrost regions, the active layer is the upper part of the ground that undergoes seasonal freezing and thawing.

With respect to rock frost weathering (see FROST AND FROST WEATHERING), alternate freezing and thawing is much more damaging than continued cold, and the effectiveness of frost is dependent on the frequency of temperature fluctuations about the freezing point in the presence of water (Ollier 1984: 125). Nevertheless, the number of freeze–thaw cycles undergone by materials unfortunately cannot be used as a direct measure of frost action effectiveness for several reasons. First, the use of air temperatures to define cycles is not satisfactory at all, since significant differences exist between air and ground temperature. This can be caused for example by the insulating effect of snow or by insolation on dark rock surfaces (Washburn 1973: 58). Second, the exact freezing temperature across which the oscillations should be measured is difficult to define, as all the water contained in soils and rocks does not freeze instantaneously, nor always at 0 °C but at negative temperatures, because, for example, of the capillary forces existing in the porous media or the supercooling phenomenon. Freezing temperature can also be lowered in presence of salts or clay. Freezing has been reported to begin at temperatures lower than −10 °C in the case of rocks characterized by very small pores.

Finally, what constitutes a freeze–thaw cycle is debatable, as some authors define specific minimum negative temperatures that have to be reached for most of the rock-absorbed water to freeze, or minimum durations for the periods at negative and positive temperatures between successive cycles. For example, according to different studies, one cycle is completed when the hourly rock temperature changes from $\geq +1$ °C to ≤ -1 °C and then back to $\geq +1$ °C (Lewkowicz 2001: 359), or when a fall below −2 °C is followed by a rise above +2 °C (Matsuoka 1991: 276). Although these thresholds have been defined in order to take into account the actual stresses undergone by the rock as accurately as possible, they make any comparison of cycle frequencies reported in different studies very difficult.

Other important components of freeze–thaw cycles with respect to frost weathering are the duration of freezing (the time period during which negative temperatures persist), the intensity of freezing (the extent of temperature decrease below 0 °C) and the rate of freezing (the rapidity or slowness with which temperature decreases below 0 °C) (McGreevy and Whalley 1982: 158). The influence of these three parameters is quite controversial.

As far as the intensity of freezing is concerned, since the greatest part of pore water freezes between 0 °C and −5 °C, volume expansion causing frost weathering of rock occurs mostly in this temperature range (McGreevy and Whalley 1982: 159; Matsuoka 1991: 272). This explains why frost decay rates do not change significantly between freeze–thaw cycles reaching minimum temperatures of −8 or −30 °C.

The impact of freezing duration has to be viewed in relation to the intensity of freezing. It is the pore sizes that determine the freezing point of water within rocks. Thus freezing occurs over a range of gradually decreasing temperatures and rocks undergo some stress only if the required critical temperatures have been reached, and for a period long enough so that the temperature change propagates from the rock surface into the centre of a block or into a rockwall. There must indeed be time for the transfer of the necessary latent heat to cause the freezing or thawing of the water in the rock (Ollier 1984: 125). The duration of the period at minimum temperature has been considered by laboratory work as completely insignificant (under constant temperature and if the freezing front stopped progressing, no breaking strain can be built up) or quite important (in an open system with a constant unfrozen water supply, segregation ice lenses may keep growing by unfrozen water migrations under constant temperature conditions). On the other hand, in field studies carried out in alpine environments where wedging (see FROST AND FROST WEATHERING) of a massive rock mass is the predominant decay process, freezing intensity and duration have been considered as fundamental parameters as they are responsible for the depth reached by the freezing front. Only long freezing periods, with stagnation of the freezing front at depths between 10 and 50 cm, are able to furnish large slabs in addition to small blocks (Coutard and Francou 1989: 415).

Various rates of freezing can favour various weathering mechanisms and lead to different degrees and types of rock decay in the same rock type. Quick cooling favours bursting and wedging effects, as more pressure is built up in pores and cracks when no time is left for water migration to occur and to relieve some of this pressure. On the contrary, slow cooling offers optimal conditions for the formation of segregation ice lenses and for scaling effects. Numerous works report higher degrees of decay after quicker frosts although some studies argue that freezing rate is not a particularly critical parameter (McGreevy and Whalley 1982: 158), or stress on the quite complex impact of freezing rates, making the evaluation of its importance difficult (Matsuoka 1991: 272). According to Matsuoka, slow freezing in an open system results in prolonged water migration toward the freezing front and, hence, in rising ice force. In contrast, in a closed system, rapid freezing favours a large ice growth strain, because pore ice contracts with time.

Rates of freezing measured in natural environments generally range between 0.2 and 4 °C per hour. However, laboratory simulation usually favours quick cooling rates (in order to accelerate COMMINUTION and the achievement of decay results) and values higher than 10 °C per hour are not uncommon. Results obtained by such experimentation may not reflect natural environmental processes.

Freeze–thaw cycles have been the subject of data collection in the field and of laboratory investigations, testing the impact of different temperature regimes on frost susceptibility. A large variety of cycle characteristics have been used, but the two main types reflect a daily moderate freezing regime (down to −8 °C) characteristic of polar maritime regions and a more intense and prolonged freezing regime (down to −30 °C) characteristic of polar continental areas.

References

Coutard, J.P. and Francou, B. (1989) Rock temperature measurements in two alpine environments, implications for frost shattering, *Arctic and Alpine Research* 21(4), 399–416.

French, H.M. (1996) *The Periglacial Environment*, Harlow: Longman.

Lewkowicz, A.G. (2001) Temperature regime of a small sandstone tor, Latitude 80°N, Ellesmere Island, Nunavut, Canada, *Permafrost and Periglacial Processes* 12, 351–366.

McGreevy, J.P. and Whalley, W.B. (1982) The geomorphic significance of rock temperature variations in cold environments: a discussion, *Arctic and Alpine Research* 14(2), 157–162.

Matsuoka, N. (1991) A model of the rate of frost shattering: application to field data from Japan, Svalbard and Antarctica, *Permafrost and Periglacial Processes* 2, 271–281.

Ollier, C. (1984) *Weathering*, Harlow: Longman.

Washburn, A.L. (1973) *Periglacial Processes and Environments*, London: Arnold.

Further reading

Lautridou, J.P. and Ozouf, J.C. (1982) Experimental frost shattering: 15 years of research at the Centre de Géomorphologie du CNRS, *Progress in Physical Geography* 6(2), 215–232.

Prick, A. (1995) Dilatometric behaviour of porous calcareous rock samples undergoing freeze–thaw cycles. Some new results, *Catena* 25(1–4), 7–20.

SEE ALSO: experimental geomorphology; frost and frost weathering; mechanical weathering; periglacial geomorphology; weathering

ANGÉLIQUE PRICK

FRINGING REEF

The morphology and genesis of CORAL REEFS varies significantly. They may be divided into ATOLL reefs, barrier reefs and fringing reefs (Nunn 1994). The youngest and most ephemeral of the three forms are fringing reefs, which also often lack the breadth, continuity and species diversity of atoll and barrier reefs. In addition, because they are located nearest the land – and indeed cannot exist distant from it – fringing reefs are those which are usually most affected by humans.

Development of fringing reefs

Unlike atoll reefs and barrier reefs, most fringing reefs formed as discrete units only during the most recent period of postglacial sea-level rise. Most began growing from shallow depths on the flanks of a tropical coastline when ocean-water temperatures (and other factors) at the end of the glacial period became suitable for reef growth. Encouraged by sea-level rise, the nascent fringing reefs began growing upwards and exist as living entities today only if they were able to 'keep up' or 'catch up' with sea level during the transgression (Neumann and MacIntyre 1985).

Once sea level reached its maximum level during the Holocene (about 5,000 cal. yr BP), 'keep-up' fringing reefs would have stopped growing mainly upwards and would have begun growing laterally, an ecological transformation involving a change in coral species distribution. Branching corals in particular would slowly have been replaced by other species adapted to outward rather than upward growth. A classic study is that of Hanauma Reef which began growing about 7,000 years ago on the inner flanks of an ancient volcanic crater on the Hawaiian island Oahu (Easton and Olson 1976).

On the other hand, 'catch-up' fringing reefs would not by definition have been able to keep pace with rising sea level and may have 'caught up' only when sea level was falling during the late Holocene. In such cases the change from upward to outward growth may have occurred more recently.

The outward growth of a fringing reef is constrained by the slope angle of the coastline from which it rises. On steeply sloping coasts, like those of the central Pacific island Niue, it is no surprise that fringing reefs are barely noticeable (and have little role in shoreline protection), often no more than a few metres in width. On coasts which slope more gently, fringing reefs may reach several hundred metres in breadth and have well-defined morphological zones (see below).

Some writers like Davis (1928) believed that a fringing reef was part of a genetic continuity and would eventually become a barrier reef and finally, when the land from the flanks of which the reef rose was submerged, an atoll reef. This is valid in only a general sense but did not take into account the effects of sea-level changes and the fact that, at the end of each Quaternary glacial period, fringing reefs re-grew. Such writers often equated the presence of fringing reefs with a coastline that had just begun sinking and, where a barrier reef was found farther offshore, would often cite a complex series of tectonic (rather than sea-level) movements to explain the association.

Morphology of fringing reefs

Fringing reefs have morphological characteristics that are shared with atoll and barrier reefs and others which are not. Along their outer, submarine slopes, fringing reefs have slopes of talus derived from the mechanical erosion of the reef edifice. Owing to the youth of fringing reefs and the comparative shallowness of the adjoining seafloor (usually a lagoon floor), these talus slopes are generally less voluminous than the equivalent features off barrier or atoll reefs. Similarly, owing to the wave energy being generally less along the fronts of fringing reefs (because waves hitting fringing reefs are commonly generated within a lagoon or are residual waves reduced in amplitude from crossing a barrier reef), reef growth and coral diversity on the outer reef crest is generally less than on barrier or atoll reefs. Yet, where a fringing reef faces directly into the ocean, these features and others are of the same size as on barrier or atoll reefs. A good example is the south coast of Tongatapu Island in the South Pacific where the south-east trade winds drive swells straight onto the narrow fringing reef which has well-developed spur-and-groove morphology along its front and an impressively high algal (*Porolithon*) ridge (Nunn and Finau 1995).

Behind the outer reef crest of fringing reefs is generally found a reef flat several tens of metres broad in which there are comparatively few living corals but an abundance of fossil reef, often planed down from a higher level. A good example is from New Caledonia (Cabioch *et al.* 1995). Particularly if the fringing reef has been significantly affected by humans (see below) the back reef area may be covered with seagrass beds or the alga *Halimeda*, sometimes terrigenous sand, all of which inhibit reef growth and may in consequence reduce the supply of calcareous sand to adjacent shorelines.

At the back of many fringing reefs is a 'boat channel' eroded in the reef surface at the point where freshwater comes out of the adjacent land. Freshwater springs are common in such places.

Emerged fringing reefs

Along those coasts where coral-reef upgrowth was able to keep pace with postglacial sea-level rise, and the sea level exceeded its present level during the middle Holocene, it is expected that fringing reefs would have grown above their present levels and that remnants of such 'emerged' fringing reefs would now be visible to testify to this. The morphology of emerged fringing reefs is often comparable to that of their modern counterparts although many are much reduced by erosion.

In the Hawaiian Islands, for example, many years of searching for emerged fringing reefs bore fruit only quite recently (Grigg and Jones 1997).

Human impact on fringing reefs

Fringing reefs are those most vulnerable to deleterious human impact. Many bear the brunt of indirect impacts like pollution and sedimentation from adjacent land areas. Direct impacts, particularly along coasts where fringing reefs are central to subsistence economies or to recreational activities, include overexploitation of edible reef organisms, trampling by humans, physical damage from boat anchors, and even poisoning or dynamiting for easy kills of large numbers of reef fish.

References

Cabioch, G., Montaggioni, L.F. and Faure, G. (1995) Holocene initiation and development of New Caledonian fringing reefs, SW Pacific, *Coral Reefs* 14, 131–140.

Davis, W.M. (1928) *The Coral Reef Problem*, Special Publication 9, Washington, DC: American Geographical Society.

Easton, W.H. and Olson, E.A. (1976) Radiocarbon profile of Hanauma Reef, Oahu, Hawaii, *Geological Society of America, Bulletin* 87, 711–719.

Grigg, R.W. and Jones, A.T. (1997) Uplift caused by lithospheric flexure in the Hawaiian Archipelago as revealed by elevated coral deposits, *Marine Geology* 141, 11–25.

Neumann, A.C. and MacIntyre, I. (1985) Reef response to sea-level rise: keep-up, catch-up or give-up, in *Proceedings of the 5th International Coral Reef Congress* 3, 105–110.

Nunn, P.D. (1994) *Oceanic Islands*, Oxford: Blackwell.

——and Finau, F.T. (1995) Late Holocene emergence history of Tongatapu island, South Pacific, *Zeitschrift für Geomorphologie* 39, 69–95.

SEE ALSO: coral reef

PATRICK D. NUNN

FROST AND FROST WEATHERING

Frost action is a collective term describing a number of distinct processes which result mainly from alternate freezing and thawing of water in pores and cracks of soil, rock and other material, usually at the ground surface. It is widely believed that frost action is the fundamental characteristic of present-day periglacial environments. Frost-action processes probably achieve their greatest intensity and importance in such areas.

In soils, FROST HEAVE, NEEDLE-ICE formation, frost creep and thermal contraction cracking are very common frost-related processes. The term *cryoturbation* refers to all soil movements due to frost action (French 1996).

Frost weathering (also called *frost shattering, congelifraction, gelifraction* or *gelivation*) contributes to the *in situ* mechanical breakdown of rocks by various processes. The conventional view is that rock decay is due to the fact that when water freezes it expands by about 9 per cent. This creates pressures, calculated to be around 2,100 kg cm^{-2} at −22 °C, that are higher than the tensile strength of rock (generally less than 250 kg cm^{-2}). However, this process rarely induces critical pressures, reached only when freezing occurs in a closed system with a very high rock moisture content (about 90 per cent). Such conditions are not common in natural environments, but when occurring, the volume expansion effect may cause rock *bursting*.

A more realistic model, also applicable to soils, is the segregation ice model (Hallet *et al.* 1991), which treats freezing in rock as closely analogous to slow freezing in fine-grained soils. When water freezes in rock or soil, the ice nuclei attract unfrozen water from the adjoining pores and capillaries. Tensions are primarily the result of these water migrations to growing ice lenses (Prick 1997). Frost weathering is induced by the progressive growth of microcracks and relatively large pores wedged open by ice growth. In the segregation ice model, low saturation in hydraulically connected pores (open system) does not preclude water migration and crack growth. The detachment of thin rock pieces by the growth of ice lenses is called *scaling*.

Frost wedging refers to rock fracturing associated with the freezing of water in existing planes of weakness, i.e. cracks and joints. Wedging can be caused by the volume expansion of water turning into ice in cracks, or by hydraulic pressure. According to this second process, the freezing front penetration in a rockwall induces a freezing of the most external part of the crack first, creating a solid plug of ice. In depth, where the saturated crack is thinner (and the freezing point thus lower), some water can be trapped under pressure by the ice growing further in from the rock surface and so contribute to crack growth outwards and downwards. In both cases, the thinner the crack, the quicker and the more severe the frost has to be in order to cause a wedging effect.

The rate at which frost shattering occurs depends on climatic factors and rock characteristics. Among climatic factors, the most important ones are the number of FREEZE–THAW CYCLES and the availability of moisture. Some thermal characteristics of the freeze–thaw cycle can also have some importance, like the freezing rate or the duration of the freezing period.

The water availability in the environment and the rock moisture content are certainly the most critical elements for defining the susceptibility of this environment to frost action (Matsuoka 1990). Laboratory experiments have shown that the amount of disintegration in rocks supplied with abundant moisture is greater than that in similar rocks containing less moisture. For this reason, dry tundra areas and cold deserts may undergo less extreme frost weathering than moister environments.

If some particular locations are characterized by a continuous and abundant water supply (for example a block sitting next to a lake shore or to a melting snow patch), a large majority of blocks exposed in cold-climate environments experience neither close to saturation conditions (because of insufficient water supply), nor a dry state (intense drying is rare).

A critical degree of saturation can be defined as a threshold moisture level for each rock type (Prick 1997): only when moisture exceeds this level will the material be damaged by frost. This parameter reflects the influence of rock characteristics on frost susceptibility and defines the part of the porous medium that has to be free of water in order not to build up a breaking strain.

The nature and characteristics of the rock are indeed a crucial factor for frost susceptibility. Rocks such as tough quartzites and igneous rocks tend to be most resistant, while porous and well-bedded sedimentary rocks, such as shales, sandstones and chalk, tend to be least resistant. Among the rock characteristics influencing frost weathering, the most determinant ones are: the rock specific surface area, permeability, porosity, pore size distribution, and mechanical strength.

A large specific surface area (i.e. internal surface of the porous media) induces a larger contact area between rock and water and therefore enhances a higher susceptibility to frost decay. A high rock permeability, by allowing easy and quick water migration, prevents critical pressures to build up (Lautridou and Ozouf 1982). Rocks with a very poor porosity are not frost susceptible: experimentation showed that rocks with a porosity of less than 6 per cent are little damaged after several hundreds of freeze–thaw cycles (Lautridou and Ozouf 1982); further research showed that this threshold value can be considered as a valuable, but rough estimate.

Pore size distribution (also called *porosimetry*) can influence frost susceptibility in various ways. Rock porous media characterized mainly by large pores (macroporosity) will tend to be frost resistant, as macroporosity favours a good permeability. Unimodal porous media (i.e. characterized by one predominant pore size) offer ideal conditions for segregation ice formation; rocks with such a pore size distribution tend to be susceptible to any type of freezing (even with a moisture content far below saturation and with a slow cooling rate) and will undergo an increased decay as freezing/thawing goes on. Multimodal porous media (i.e. characterized by pores of various sizes) are not favourable to the set up of large-scale water migrations; rocks with such a pore size distribution tend to be frost susceptible only with high moisture content, preferably in the case of a quick freezing.

Among ROCK MASS STRENGTH parameters, tensile strength has a considerable influence on rock frost decay (Matsuoka 1990). Crack density and width often influence water penetration in the bedrock and allow wedging to take place.

Frost action is one component of *cryogenic weathering*, i.e. the combination of weathering processes, both physical and chemical that operate in cold environments either independently or in combination. Many aspects of cryogenic weathering are not fully understood, but the processes other than frost that may be efficient decay agents are: HYDRATION (see WETTING AND DRYING WEATHERING), thermal fatigue (see INSOLATION WEATHERING), SALT WEATHERING, CHEMICAL WEATHERING, ORGANIC WEATHERING and PRESSURE RELEASE (particularly in recently deglaciated areas). Solutional effects are present in limestone and KARST terrain exists in PERMAFROST regions. The dominance of frost action among these processes is considered as doubtful, but the definition of the exact role of each of these processes in the different cold environments and in the different periods of the year is problematic.

Frost weathering characteristically produces angular fragments of various sizes. In periglacial areas, cryogenic weathering determines the formation of some extensive features like blockfields (see BLOCKFIELD AND BLOCKSTREAM), GRÈZE

LITÉES, SCREES, TALUS slopes or ROCK GLACIERS. Its action is also often crucial for MASS MOVEMENT processes like rock avalanches and rock falls.

The predominant size to which rocks can be ultimately reduced by frost action is generally thought to be silty particles with grain sizes between 0.01 and 0.05 mm in diameter. Experimentation on mineral particles indicated that frost weathering occurs within the layer of unfrozen water adsorbed on the surface of these particles. The minerals' susceptibility to weathering depends not so much on their mechanical strength as on the thickness and properties of this unfrozen water film. Decay occurs when this water film becomes thinner than the dimensions of the microcracks and defects that characterize the surface of mineral particles. The protective role of the stable film of unfrozen water is highest with silicates, such as biotite and muscovite, and lowest with quartz. Experimentation results indicate that under cold conditions the ultimate size reduction of quartz (0.01–0.05 mm) is smaller than for feldspar (0.1–0.5 mm), a reversal of what is assumed for temperate or warm environments (Konishev and Rogov 1993).

Frost weathering is studied both in the field and in the laboratory. The most commonly used techniques are: visual observation and photographic documentation of the decay evolution, weight loss, frost-shattered debris characterization, assessment of mechanical properties (like tensile or compressive strength) or elasticity properties (Young's modulus), ultrasonic testing, evolution of porosity and pore size distribution, dilatometry, and crack opening assessment.

Some field studies have been undertaken with the aim of increasing the availability of data upon rock temperature and moisture content. The lack of such data has up to now been a considerable impediment to a definition of the exact role of frost action in cryogenic weathering and to the realization of laboratory simulation using thermal and moisture regimes likely to occur in natural environments. Other studies focus on the rate of bedrock weathering by frost action (Lautridou and Ozouf 1982) and on the definition of predictive models (Matsuoka 1990). Laboratory simulation and modelling identify the climatic conditions and rock characteristics that emphasize frost action efficiency and so define the exact role of the various weathering mechanisms (e.g. Hallet *et al.* 1991; Prick 1997).

A major gap remains between field and laboratory research (Matsuoka 2001). This is due to a difference in the size of the study object (small blocks in the laboratory, but rockwalls sometimes in the field), in the type of rock material (intact soft rocks with medium or high porosity are overrepresented in laboratory simulations, but jointed massive rocks with low porosity are very common in cold environments) and thus in the type of frost weathering process taken into account (mostly bursting or scaling in the laboratory simulations; mostly wedging in the field). Wedging may sometimes be the only frost weathering process acting on fractured rock characterized by a low porosity and a high mechanical strength for the individual blocks. This may lead to *macrogelivation*, i.e. frost weathering at a large scale, acting mainly through the crack system, as opposed to *microgelivation*, which refers to frost decay acting in the porous media of individual small-sized blocks. This further illustrates the inadequacy of a simplistic view of frost weathering.

References

French, H.M. (1996) *The Periglacial Environment*, Harlow: Longman.

Hallet, B., Walder, J.S. and Stubbs, C.W. (1991) Weathering by segregation ice growth in microcracks at sustained subzero temperatures: verification from an experimental study using acoustic emissions, *Permafrost and Periglacial Processes* 2, 283–300.

Konishev, N.V. and Rogov, V.V. (1993) Investigations of cryogenic weathering in Europe and Northern Asia, *Permafrost and Periglacial Processes* 4, 49–64.

Lautridou, J.P. and Ozouf, J.C. (1982) Experimental frost shattering: 15 years of research at the Centre de Géomorphologie du CNRS, *Progress in Physical Geography* 6(2), 215–232.

Matsuoka, N. (1990) The rate of bedrock weathering by frost action: field measurements and a predictive model, *Earth Surface Processes and Landforms* 15, 73–90.

——(2001) Microgelivation versus macrogelivation: towards bridging the gap between laboratory and field frost weathering, *Permafrost and Periglacial Processes* 12, 299–313.

Prick, A. (1997) Critical degree of saturation as a threshold moisture level in frost weathering of limestones, *Permafrost and Periglacial Processes* 8, 91–99.

Further reading

White, S.E. (1976) Is frost action really only hydration shattering? A review, *Arctic and Alpine Research* 8(1), 1–6.

Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

SEE ALSO: experimental geomorphology; mechanical weathering; periglacial geomorphology; weathering

ANGÉLIQUE PRICK

FROST HEAVE

Frost heave is best known from the wintertime uplift of the ground surface, familiar to dwellers in cold climates, which is evidenced by jammed gateways, uneven roads, cracked foundations and the breaking-up of road surfaces in the spring thaw. These effects are not ascribable to the expansion of water that occurs on its freezing (9 per cent). They are due to the movement of water into the soil that is freezing, with the formation of accumulations of ice – increasing the soil volume, giving displacement (the 'heave'). These ice structures are called 'lenses', 'schlieren' or ice 'masses' and known collectively as segregation ice (because each is larger than pore size and has been segregated from the soil pore structure). Segregation ice is not ice from entrapped snow, buried glacial ice or buried lake or marine ice, although it may reach cubic metres in size. Its nature and extent depends on the nature of the granular soil material and a variety of local factors (drainage conditions, temperature regime, depth in the ground, etc.). Thus frost heave is commonly uneven, giving rise to irregularities of the ground surface (bumpiness) usually recurring year after year or, in PERMAFROST, persisting for many years. The forces generated by the heaving material can be very large.

Segregation ice and thus frost heave is an expression of the fundamental thermodynamic behaviour of a porous medium on freezing; this thermodynamic behaviour is ultimately responsible for the main properties and characteristics associated with soils in cold climates. As a consequence frost heave has enormous economic (geotechnical) significance; overcoming its effects is the essential problem for construction of buildings, roads, airports and pipelines in the cold regions.

The processes associated with frost heave largely explain the origin of most terrain forms occurring naturally and characteristically in cold climates – so-called 'periglacial' (see PERIGLACIAL GEOMORPHOLOGY) features. Boulders ('growing stones') are heaved to the ground surface by annual cycles of freezing and rearrangement of soil particles at thawing. Incremental frost heave is an important process in formation of PATTERNED GROUND, such as stone circles and stone polygons, where stones and boulders are heaved in particular directions (as a function of temperature and other factors) to give rise to conspicuously ordered surface arrangements. PINGOS, features occurring locally in regions with permafrost, look like volcanic cones and are elevated by the large, hidden central core of ice. They are the product of a particular thermal regime, commonly involving the gradual freezing of previously unfrozen ground below a receding water body. The frost heave process is largely responsible for lifting the above-surface material in pingos to elevations of tens of metres, so the forces developed must be large.

The instability of slopes, and the development of certain forms of SOLIFLUCTION, mudflows and landslides are ascribable to the excess water released on thawing of frost-heaved soils with their ice segregations, and the associated high PORE-WATER PRESSURES. Not infrequently the volume of segregated ice exceeds the volume of water the soil can hold in the thawed state by a factor of two or more. This accounts for a greatly weakened state of the newly thawed soil.

Fundamentally, the water moves toward a zone of freezing in the soil because of thermodynamic potentials arising with the growth of ice crystals in small spaces. Although the thermodynamic principles have been recognized for more than a century (and also describe, for example, crystallization phenomena in solutions, the formation of ice crystals or of water droplets in the atmosphere, or the nucleation of bubbles in liquids) the significance for soils has been realized fully only in recent decades. The thermodynamic potential may be regarded, with some simplification, as a pressure, and is referred to by different terms in different branches of science and technology. The pressure of the water falls with temperature in freezing soil, so that there is a gradient from warm (unfrozen) to cold (frozen). However, the pressure of the ice in the ice segregations rises as the temperature falls, and it is demonstrably this pressure which causes frost heave. Furthermore, the thermodynamic relations require ice to form in larger spaces and pores first. As a consequence, there is an unfrozen water content of frozen soils, decreasing with temperature as progressively smaller pores are filled by ice, and which is, therefore, a function of pore size distribution.

The significance of the soil water accumulating (the process of frost heave) and then freezing in this way over a range of temperatures down to several degrees below 0 °C, is that the thermal and mechanical properties of the frozen soils are highly temperature dependent. Frozen, heaved soils are prone to creep in a manner rather similar to glaciers but with lower rates of deformation; this is probably the cause of certain large vegetation-covered solifluction terraces on slight slopes. The grain-size and pore size distribution of a soil are crucial to its behaviour when frozen because they control the (unfrozen) water content of the specific soil. The release of latent heat of freezing of the water effectively controls the heat capacity of the soil; thermal conductivity is also modified (though less so) because of the difference in thermal conductivity between ice and water. The thermal diffusivity, which is the ratio of thermal conductivity to heat capacity, is consequently highly temperature dependent and controlled by the pore size distribution – that is, by the nature of the soil (clay, silt or sand, or combinations of particle sizes), and the amount of frost heave.

The thermal diffusivity controls such phenomena as the depth to which winter freezing occurs, and the depth of summer thawing (the ACTIVE LAYER) above permafrost. Indeed the distribution of permafrost itself (ground remaining frozen year in, year out), in depth and in time (and in response to climate or microclimate change), depends substantially on its thermal diffusivity. Terrain features, ascribable to frost heave and associated with the comings and goings of permafrost, include ALASes, palsa and THERMOKARST.

Counteracting effects of frost heave and subsequent thawing added billions of dollars to the cost of the transAlaska oil pipeline. The forces generated (CRYOSTATIC PRESSURES) by frost heave around gas pipelines in permafrost regions threaten their stability and thus their financial viability. Avoidance of frost heave is the main reason for added costs of infrastructure in general in the cold regions; these added costs are greatest in the 25 per cent of the Earth's land surface underlain by permafrost but are also a major factor in construction (especially of highways and airports) in the further 20 per cent or so which has significant winter frost penetration, and consists largely of highly populated temperate lands.

When Taber (1918) first clearly demonstrated that the geotechnical problem of frost heave was due to the migration of water with accumulation of excess ice in the frozen ground, he paved the way for Beskow's classic work (1935) on frost heave and its significance in relation to the local environment (soil type, groundwater conditions, confining pressures, etc.). In 1943 the remarkable study by Edlefsen and Andersen (resulting from the wartime collaboration of two scientists in different fields) established the thermodynamic interpretation, which substantiates the largely empirical approach that has been used by geotechnical specialists concerned with engineering (Andersland and Ladanyi 1995) for cold regions development in the broadest sense. Agronomists too, have an important involvement. Today, geocryology, the study of the ground surface regions in freezing climates (Williams and Smith 1989) notably developed in Russia (Yershov 1998), is concerned mainly with the effects of frost heave, a phenomenon first recognized some two hundred and fifty years ago (see Beskow 1935).

References

Andersland, O.B. and Ladanyi, B. (1995) *An Introduction to Frozen Ground Engineering*, London: Chapman and Hall.

Beskow, G. (1935) Tjällyftningen och tjällyftningen med særskild hensyn til vägar och järnvägar. *Sveriges Geologiska Undersökning. Avh. och Uppsats., Ser. C, 375, årsbok 26, 3.* (Available in translation: Soil freezing and frost heave with special application to roads and railroads, in P.B. Black and M.J. Hardenberg (eds) (1991) Historical perspectives in frost heave research, *US Army, Cold Reg. Res. and Engg. Labs*, Special Report 91–23, Hanover, NH.

Taber, S. (1918) Surface heaving caused by segregation of water forming ice crystals, *Engineering News Record* 81, 683–684.

Williams, P.J. and Smith, M.W. (1989) *The Frozen Earth. Fundamentals of Geocryology*, Cambridge: Cambridge University Press.

Yershov, E.D. (ed. P.J. Williams) (1998) *General Geocryology* (translated from: Obschaya geokriologiya, Nedra 1990), Cambridge: Cambridge University Press.

PETER J. WILLIAMS

G

GENDARME

A needle-shaped rock pinnacle located along a mountain ridge or arête. The term gendarme is universal, yet employed predominantly in alpine geomorphology and mountaineering. Gendarme shares its name with a French policeman, as both may block one's passage and hinder progress. They are commonly found in the Alps, such as Pic de Roc gendarme in Chamonix, French Alps. However, similar forms exist in other mountainous regions, such as Bryce Canyon, USA. Gendarmes are also referred to as rock pinnacles and aiguilles, yet are generally more pointed and larger than an aiguille.

SEE ALSO: rock and earth pinnacle and pillar

STEVE WARD

GEOCRYOLOGY

The study of Earth materials having a temperature below 0 °C. Washburn (1979) recognized that it was sometimes taken to include glaciers, but argued that it was more specifically used as a term for PERIGLACIAL GEOMORPHOLOGY and PERMAFROST phenomena. Indeed, the subtitle of his magisterial volume *Geocryology* was 'A survey of periglacial processes and environments'. In that volume he studied such phenomena as frozen ground, FROST AND FROST WEATHERING, PATTERNED GROUND, avalanches (see AVALANCHE, SNOW), SOLIFLUCTION, NIVATION and THERMOKARST.

Reference

Washburn, A.L. (1979) *Geocryology: A Survey of Periglacial Processes and Environments*, London: Edward Arnold.

A.S. GOUDIE

GEODIVERSITY

The geodiversity concept first appeared in Australia (especially Tasmania), and received wider recognition, even if always not proper understanding, in the mid-1990s. This robust geodiversity concept has been poorly developed yet in methodological terms. The most popular definition of geodiversity was put forward by the Australian Natural Heritage Charter (AHC 2002):

> Geodiversity means the natural range (diversity) of geological (bedrock), geomorphological (landform) and soil features, assemblages, systems and processes. Geodiversity includes evidence of the past life, ecosystems and environments in the history of the earth as well as a range of atmospheric, hydrological and biological processes currently acting on rocks, landforms and soils.

Geodiversity is now being used in a very holistic way to emphasize the links between geosciences, wildlife and people in one environment or system. The above definition can be supplemented with the statement that geodiversity also embraces quantitative and qualitative topics or indicators at any timescale which make it possible to distinguish marked peculiarities of a georegion, a spatial unit of an unspecified taxonomic rank. This means that bedrock, landforms and soils can be classified by at least two important categories: uniqueness and representativeness. From the geomorphological diversity perspective, an outstanding landform is a feature which is rare, unique, an exceptionally well-expressed example of its kind, or otherwise of special importance within a georegion. A representative landform, in turn, may be either rare or common, but is considered significant as a well-developed or well-exposed example of its kind. A landform type or system can be characterized by an isotropic entity in terms of topographic shape, physical contents, morphogenetic controls and processes, as well as time of formation.

The term geodiversity is commonly used in two meanings, simpler and broader. The first refers to the total range, or diversity, of geological, geomorphic and soil phenomena, and treats geodiversity as an objective, value-neutral property of a real geosystem. In this case a statement of the diversity is made, but the geosystem is not assessed in terms of what kind of geodiversity it is: low or high? The other usage conveys the idea that geodiversity refers specifically to particular geosystems that are in themselves diverse or complex, and thus does not apply to systems which are uniform or have low internal diversity. An example can be the valley of a river flowing through mountains, uplands and lowlands, filled with a wealth of valley, channel and bedforms, and therefore showing high geodiversity, whereas an area of lowland without any streams, basins and/or hummocks has low geodiversity. Questions about the measure of geodiversity are troublesome. Which area displays higher geodiversity: one in which there are 15 mogotes, or another featuring 5 volcanoes, 5 glaciers and 5 river valleys? Or another: has geodiversity increased or diminished in an area transformed by numerous and extensive man-made changes? Landforms are defined by their surface contours and that is why some people claim that the disturbance of significant landform contours (e.g. by excavation) will by definition degrade their geodiversity values, while others see this morphological disturbance as enrichment of geodiversity. Obviously, this situation calls for some clear-cut criteria of geodiversity. One of the possible solutions is a hierarchical classification of landforms: morphoclimatic zone (polar), morphogenetic zone (mountain), morphosystem (glacial system), type of relief (depositional relief), set of landforms (morainic landforms), and single form (terminal moraine). This classification is a function of complexity (see COMPLEXITY IN GEOMORPHOLOGY) reduction. One might argue that an increase in complexity entails an increase in geodiversity, and variations in this relationship are a matter of two functions: asymptotic and exponential.

Because geodiversity is valuable from a variety of perspectives (intrinsic, ecological, geoheritage, as well as scientific, educational, social, cultural, tourist, etc.) it should undergo geoconservation as a result of which it is possible to create GEOSITES for present and future generations.

It should be added that the term geodiversity is analogous to the term biodiversity, which is used to denote species, genetic and ecosystem diversity. It is important to note that the only analogy is that both involve a diversity of phenomena and beyond this self-evident similarity, no further analogies between the nature of ecological and geomorphic processes are expressed or implied. For example, both processes contrast strongly in their timescales. Ecosystems with plant or animal life cycles of tens to hundreds of years do not closely parallel the much longer term active or relict geosystems with weathering, erosion and sedimentation, or Earth internal processes such as seismic or volcanic activity and plate tectonics controlled by processes acting over many thousands or millions of years.

Reference

AHC (2002) *Australian Natural Heritage Charter for the Conservation of Places of Natural Heritage Significance*, Australian Heritage Commission in association with Australian Committee for IUCN, Sydney.

ZBIGNIEW ZWOLINSKI

GEOINDICATOR

The concept of geoindicators was put forward by the International Union of Geological Sciences in 1992. The task of the IUGS working group was to draw up an inventory of indicators to be measured and evaluated under any programme of abiotic environment monitoring. The inventory is not supposed to be a universal standard, but rather to provide a list for the selection by environment monitoring teams of those indicators that can be usefully employed with reference to their study area and time period. Thus, while the list of twenty-seven geoindicators is finite, their choice for the description of environmental change is free. Each geoindicator was evaluated relative to a set of checklist parameters: name, description, significance, human/natural cause, applicable environment, types of monitoring sites, spatial scale, method of measurement, frequency of measurement, limitations of data, applications to past and future, possible thresholds, key references, other information sources, related issues and overall assessment (see Table 20).

Table 20 Geoindicators: natural* vs. human influences**, and utility for reconstructing past environments***

Geoindicator	N*	H**	P***
Coral chemistry and growth patterns	High	High	High
Desert surface crusts and fissures	High	Moderate	Low
Dune formation and reactivation	High	Moderate	Moderate
Dust storm magnitude, duration and frequency	High	Moderate	Moderate
Frozen ground activity	High	Moderate	High
Glacier fluctuations	High	Low	High
Groundwater quality	Moderate	High	Low
Groundwater chemistry in the unsaturated zone	High	High	High
Groundwater level	Moderate	High	Low
Karst activity	High	Moderate	High
Lake levels and salinity	High	High	Moderate
Relative sea level	High	Moderate	High
Sediment sequence and composition	High	High	High
Seismicity	High	Moderate	Low
Shoreline position	High	High	High
Slope failure (landslides)	High	High	Moderate
Soil and sediment erosion	High	High	Moderate
Soil quality	Moderate	High	High
Streamflow	High	High	Low
Stream channel morphology	High	High	Low
Stream sediment storage and load	High	High	Moderate
Subsurface temperature regime	High	Moderate	High
Surface displacement	High	Moderate	Moderate
Surface water quality	High	High	Low
Volcanic unrest	High	Low	High
Wetlands extent, structure and hydrology	High	High	High
Wind erosion	High	Moderate	Moderate

Source: After ITC (1995)

From the point of view of geomorphology, especially dynamic geomorphology, the geoindicator concept seems to be particularly well-suited to determine changes in morphogenetic and sedimentary environments or, broadly speaking, in geosystems. Just like systems theory or allometric analysis, the geoindicator concept has also been adapted from biological sciences. Geoindicators are measures of surface and near-surface geological processes and phenomena that tend to change significantly in less than a hundred years, and which supply crucial information for estimating the state of the environment. This definition specifies the time interval concerned as under a hundred years, which means that geoindicators embrace those processes and phenomena that are highly variable at a short timescale. Hence geoindication will not cover processes involving slow change, like metamorphism or large-scale sedimentation. Geoindicators should answer such questions as, e.g.:

- How often does a process occur?
- What is the rate of river load transport?
- How stable is an individual landform?
- Is the given landform still active, or is it a remnant of an earlier developmental stage?

This way of question formulation determines the specific character of geoindicators: they can express the magnitude, frequency, and rate and/or behaviour trend of an event, process or phenomenon. This means that geoindicators can have widespread application in present-day geomorphological research and, when backed up by paleoenvironmental research, they can provide an excellent basis for forecasting studies. It is especially important when one considers the last decades with their climate change and the consequences it has for the operation of most geosystems throughout the globe. This characterization of geoindicators can be extended to include interactions between the abiotic and biotic environments as well as the fact that it is possible to use geoindicators for different-sized areas to measure extreme, secular and predominant events and to observe natural and man-made processes. Altogether, geomorphologists will find that they have acquired a research tool which is bound to bring about methodological changes in their field.

Reference

ITC (1995) *Tools for assessing rapid environmental changes. The 1995 geoindicator checklist*, International Institute for Aerospace Survey and Earth Sciences, Enschede, Publication Number 46.

Further reading

Berger, A.R. and Iams, W.J. (eds) (1996) *Geoindicators. Assessing Rapid Environmental Changes in Earth Sciences*, Rotterdam: A.A. Balkema.

ZBIGNIEW ZWOLINSKI

GEOMORPHIC EVOLUTION

Geomorphic evolution at its simplest means the mode of change of landform or geomorphic system over time. Qualitative theories continue to dominate geomorphology but a quantitative theory of landform evolution is becoming a central challenge. Traditional qualitative models of landform evolution include the geographical cycle (Davis 1899), Penckian morphological analysis (Penck 1924), the semi-arid erosion cycle (King 1962) and climatogenetic geomorphology (Büdel 1977). These four models represent the framework and options within which landscape evolution were considered from about 1890 until the 1960s. Each of them (except for Penckian morphological analysis) is still in vogue among those who are interested in landscape evolution at the regional scale. The geographical cycle (Davis 1899) is still widely celebrated as a uniquely effective pedagogic device. The orderly evolution of landscape through the stages of youth, maturity and old age, and its interruption at widely separate points in time by massive tectonic uplift, is intuitively appealing. Davis claimed that his model embraced the five factors of structure, process, stage, relief and texture of dissection, but much of the literature says that he considered only the first three. The single major problem with this model was the complete absence of field measurements to confirm or reject assumptions in the model. Nevertheless, there are few better models available to interpret, in a qualitative way, the massive erosional unconformities of, for example, the Grand Canyon of the Colorado River. The geographical cycle has never been proven wrong, but it has been bypassed rather than replaced.

Lester King's subaerial cycle of erosion (King 1962) is perhaps the only serious competitor with the geographical cycle as an interpreter of large-scale, low gradient erosion surfaces. King, strongly influenced by his observations on African escarpments and plateau surfaces, framed his model around the notion of the parallel retreat of scarps. He also attempted, with debatable success, to link his model with the global plate tectonics framework that was evolving during his productive career. His concept of CYMATOGENY (arching of extensive land surfaces with little rock deformation) was a necessary addition to the traditional concepts of orogeny and epeirogeny, and flies in the face of conventional plate tectonics, where massive horizontal movements are favoured. His attempts (King 1962) to correlate pre-Tertiary erosion surfaces globally have met with little debate (except for the critique of cymatogeny) because so few geomorphologists are working at this scale.

A third interesting model of geomorphic evolution is provided by Julius Büdel (1977) and known as the climatogenetic model. The major elements of this model have been interpreted for English readers by Hanna Bremer. The underlying premise is that landscapes are composed of several RELIEF GENERATIONS and the challenge is to recognize, order and distinguish these relief generations It is unfortunate that the major references in the English literature have been sceptical of the model and have failed to give a balanced review (Bremer 1984). Twidale (1976) provides a refreshing summary of the relevance of etchplains (see ETCHING, ETCHPLAIN AND ETCHPLANATION) in Australian landscape evolution.

The Penckian model (Penck 1924) was called morphological analysis. The underlying premise of his analysis is that the rate of uplift, and variations in that rate of uplift over time, dictate landform evolution. His ideas were not taken seriously in Germany, but they were widely promulgated in Anglo-America because of Davis's interest and opposition to the model. Details of the slope processes discussed are hard to verify and understand because of the lack of field data. But in its championing of endogenic processes and its time independent emphasis, this model was strongly differentiated from the first three. Time-independent models (in which the idea of evolution sits uncomfortably) have been promoted by G.K. Gilbert (1877) and J.T. Hack (1960).

The dichotomy between historical evolutionary studies and functional geomorphology implies that these two approaches do not fit easily together. Indeed, Bremer has said that geomorphology is developing along two lines: the origin of landforms is primarily being studied in continental Europe with climatogenic or tectogenic causes in the foreground. In the English-speaking world the study of geomorphic processes prevails.

Discussions by Schumm (1973), Twidale (1976), Brunsden (1980; 1993) and Ollier (1991) have attempted to reconcile these apparently contradictory positions within a largely qualitative dialogue. The essential contributions to this more recent discussion are the concept of geomorphic thresholds and complex geomorphic response (Schumm), formative events, relaxation time and landform persistence (Brunsden), the understanding that pre-Tertiary landscapes are still decipherable (Twidale), the importance of reconciling plate tectonic theory and morphological evidence (Ollier) and the disequilibrium of all landscapes influenced by Quaternary glaciation (Church and Slaymaker 1989).

A quantitative theory of landform evolution, by contrast with the theories discussed above, requires that the storage and flux rates of water, its flow paths and pressure fields be quantitatively related to their controls and that the boundary conditions of climate, rock properties, topography and stratigraphy be known. But by far the bulk of research on geomorphic evolution has taken place at meso- and micro-scales. And this is where the basic disjuncture in geomorphic thinking has been most evident. Systems modelling and mathematical modelling has tended to drive geomorphic discussion towards the smaller scale landforms, and geomorphic evolution has become, for example, slope evolution, or channel evolution or shoreline evolution.

The work of Ahnert (1967 et seq.) and Kirkby (1971 et seq.) is instructive in that they have been able to satisfy the requirements of quantitative theory by limiting the scale of their models and establishing precise boundary conditions to simulate real world slopes and basins. From 1967 to 1977, Frank Ahnert developed a series of models that used empirical equations to deal with possible ways of relating waste production, delivery and removal at a point on a slope. His final model was a three-dimensional process-response model of landform development. From 1971 to the present, Kirkby has developed increasingly integrated models of slope and drainage basin development, many of them using differential equations that constrain mass balance and thereby maintain continuity These models have had difficulty in dealing realistically with such phenomena as landslides (too rapid) and storage accumulation (too slow), but they represent the cutting edge of modelling in geomorphology from a slope process perspective.

Hydrogeomorphologists, such as Dunne, Dietrich, Montgomery and Church, have led the movement from micro-scale modelling of fluvial process towards a meso-scale modelling of drainage basins, in which they couple slope and channel processes and exploit the drainage network properties to produce more realistic dynamic drainage basin models. Howard (1994) poses a series of critical questions around the landscape modelling project. What is the simplest mathematical model that will simulate morphologically realistic landscapes? What are the effects of initial conditions and inheritance on basin form and evolution? What are the relative roles of deterministic and random processes in basin evolution? Do processes and forms in the drainage basin embody principles of optimization and, if so, why? Is there some basin characteristic form that is invariant in time even under a change in the relative role of the chief land-forming processes. The development of drainage basins requires at least two superimposed processes. He called them soil creep and water flow; in the language of the modellers, one must be a diffusional creep-like mass wasting process capable of eroding the land surface even for vanishingly small contributing areas. Such a process requires an increase in gradient downslope because of its loss of efficiency with increasing area.

The other is an advective fluvial process that increases in efficiency with increasing contributing area. The interplay of these processes produces a combination of convex and concave landforms. By enforcing continuity of flow and continuity of sediment through a coupled system of partial differential equations, the rate of change of elevation can be made dependent on the net flux of sediments as forced by linear increase in discharge. This fundamental step in the understanding of the self-organization of landscape depends on the coupling of the developing landscape with flow rate.

Willgoose *et al.* (1992) presented a catchment evolution model that was essentially a process-response model sensitive to the erosional development of river basins and their channel networks. The model describes the long-term changes in elevation with time that occur in a drainage basin as a result of large-scale mass transport processes. The mass transport processes modelled are tectonic uplift, fluvial erosion, creep, rain splash and landslides. Individual landslides are not modelled but the aggregate effect of many landslides is. The model explicitly differentiates between the part of the basin that is channel and the part that is hillslope. A channel initiation function provided by Dietrich (Dietrich *et al.* 1992) defines a threshold beyond which a channel is formed.

Both dynamic equilibrium and transient states can be modelled in this way. Howard (1994) has noted that the erosion, transport and depositional processes, especially in the river channels, have been greatly oversimplified in the Willgoose *et al.* model and he has generated both alluvial and non-alluvial channel versions of his own model. A more fundamental criticism is that the model does not clarify the linkages of fundamental aspects of the dynamics and the existence of general scaling relations in the network and the landscape itself. Hence the search for improved understanding through analysis of the fractal characteristics of river basins, particularly scale invariance, self-similarity and self-affinity. Multifractality has become a valuable property to identify changing domains of specific process sets (Montgomery and Dietrich 1994).

Understanding of the variety of modes of geomorphic evolution at a variety of spatial and temporal scales is the best evidence of progress in the field. For a number of years at the beginning of this century, researchers were expected to adopt a single model and to stick with it. As a result, the field stagnated under the influence of a single paradigm. In the contemporary state of geomorphology, one of the large issues within models of evolution at the site and basin scale relates to the relation between deterministic and probabilistic modelling.

References

Ahnert, F. (1967) The role of the equilibrium concept in the interpretation of landforms of fluvial erosion and deposition, in P. Macar (ed.) *L'Evolution des Versants*, 23–41, Liège: University of Liège.

Bremer, H. (1984) Twenty one years of German geomorphology, *Earth Surface Processes and Landforms* 9, 281–287.

Brunsden, D. (1980) Applicable models of long term landform evolution, *Zeitschrift für Geomorphologie Supplementband* 36, 16–26.

——(1993) Barriers to geomorphological change, in D.S.G. Thomas and R.J. Allison (eds) *Landscape Sensitivity*, 7–12, Chichester: Wiley.

Büdel, J. (1977) *Klima-Geomorphologie*, Berlin, Stuttgart: Borntraeger.

Church, M. and Slaymaker, O. (1989) Disequilibrium of Holocene sediment yield in glaciated British Columbia, *Nature* 337, 452–454.

Davis, W.M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.

Dietrich, W., Wilson, C.J., Montgomery, D.R., McKean, J. and Bauer, R. (1992) Erosion thresholds and land surface morphology, *Geology* 20, 675–679.

Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*. US Geographical and Geological Survey of the Rocky Mountain Region. Washington, DC: US Government Printing Office.

Hack, J.T. (1960) Interpretation of erosional topography in humid temperate regions, *American Journal of Science* 258A, 80–97.

Howard, A.D. (1994) A detachment limited model of drainage basin evolution, *Water Resources Research* 30, 2,261–2,285.

King, L.C. (1962) *The Morphology of the Earth*, Edinburgh: Oliver and Boyd.

Kirkby, M.J. (1971) Hill slope process response models based on the continuity equation, in D. Brunsden (ed.) *Slopes: Form and Process*, 15–30, Institute of British Geographers Special Publication 3.

Montgomery, D.R. and Dietrich, W.E. (1994) Landscape dissection and drainage slope thresholds, in M.J. Kirkby (ed.) *Process Models and Theoretical Geomorphology*, 221–246, New York: Wiley.

Ollier, C.D. (1991) *Ancient Landforms*, London: Belhaven.

Penck, W.D. (1924) *Die Morphologische Analyse: Ein Kapital der Physikalischen Geologie*, Geographische Abhandlungen, 2 Reihe, Heft 2. Stuttgart: Engelhorn.

Schumm, S.A. (1973) Geomorphic thresholds and complex responses of drainage systems, in M. Morisawa (ed.) *Fluvial Geomorphology*, 299–310, Binghamton: Publications in Geomorphology.

Twidale, C.R. (1976) On the survival of paleoforms, *American Journal of Science* 276, 77–95.

Willgoose, G.R., Bras, R.L. and Rodriguez-Iturbe, I. (1992) The relationship between catchment and hill slope properties: implications of a catchment evolution model, *Geomorphology* 5, 21–38.

SEE ALSO: dynamic geomorphology; fractal; geomorphology

OLAV SLAYMAKER

GEOMORPHOLOGICAL HAZARD

A significant practical contribution of geomorphology is the identification of stable landforms and sites with a low probability of catastrophic or progressive involvement with natural or man-induced processes adverse to human occupance or use. Hazards exist when landscape developing processes conflict with human activity, often with catastrophic results. People are killed and property is destroyed or damaged by extreme geomorphic events, and the toll has become greater as human activity has stretched to areas that were avoided in the past. As the population of the Earth has more than doubled from the three billion of 1960, annual losses due to disasters have grown more than ten fold (Bruce 1993).

Tragic examples abound: in 1970 a cyclonic storm surge pushed three to five metres of water into the low deltas entering the Bay of Bengal. The surge, and riverine flooding caused by discharge blocking, resulted in the deaths of an estimated 300,000 to 500,000 people in Bangladesh; an earthquake-induced debris flow descending the flanks of Mt. Huarascan that same year buried over 25,000 people in Peru; a 1991 storm-induced mudflow overwhelmed a concrete drainage channel, killing an estimated 7,000 people in the Philippines; and despite more than fifty years of comprehensive flood control, the 1993 floods along the Mississippi River were the costliest in American history.

Geomorphologists are increasingly engaged in the mapping and modelling of geophysical, hydrological and surficial material characteristics which expose areas to rupture, failure, fire, inundation, drought, erosion or submergence. Coupled with land use and human infrastructure analyses they examine the location, value, exposure and vulnerability of the human environment to hazard damage (see ENVIRONMENTAL GEOMORPHOLOGY). When the population density and demographics are added, potential casualties and emergency services needs may be forecast. This requires the integration of social scientists who focus upon social, technical, administrative, political, legal and economic forces which structure a society's strategies and policies for risk management (i.e. prevention, mitigation, preparedness, prediction and warning, and recovery), public awareness, emergency training, regulation and social insurance. Such a comprehensive approach would have been nearly inconceivable in the past, but with the advent of computerized geographic information systems (GIS) such mapping, modelling, and decision support systems are becoming more commonplace (Carrara and Guzzetti 1995).

Perplexing questions centre upon the apparent increase in frequency of catastrophic geomorphic hazards. Accurate statistical analyses of such infrequent occurrences require an observation period of well in excess of a century (Berz 1993), while consistent reporting of most types of disasters have a much shorter history. Monitoring techniques and measurement scales (e.g. Richter, Beaufort), remote sensing, and communications have only recently allowed the global reporting of events in comparable terms. A study of volcanic eruptions by Simpkin *et al.* (1981) concludes that the reported increase in volcanic activity over the past 120 years is almost certainly due to improved reporting and communications technology; they even report a reduction in 'apparent' activity during the two world wars. Despite the growing influence of global databases, scientific consortia, and the news media, additional factors may be influencing the growing number of reported disasters.

While considerable scientific debate lingers around the issue of global climatic change, geomorphologists are well aware of other indirect effects of human activity (Rosenfeld 1994b). Certainly the deforestation of large areas has caused landslides and increased both the frequency and peak flows of flood events in many areas, while overgrazing has accelerated drought effects and erosion. Groundwater withdrawal and irrigation diversions have affected natural vegetation and micro-climates of some regions, and have even induced earthquakes in some cases.

As climate models become increasingly realistic, their mathematical results consistently point to a more hazardous world in the future. That increasing concentrations of greenhouse gases in the atmosphere, primarily resulting from the burning of fossil fuels, is changing the radiation balance and perhaps the climate is consistent with recent disaster experience. There is general agreement among atmospheric scientists that a 'warmer' world would be a 'wetter' world, with no increase in the number of days with rain, but with more intense rainfall. Combined with the hydrologic effects of land use changes, the frequency and severity of floods would surely increase, especially in the monsoon climates of south Asia where flooding already reaches catastrophic proportions. Drought effects in sub-Saharan Africa, South America and Australasia could occur more frequently and be more severe as a result of intensified El Niño–Southern Oscillation events. Resultant sea-level rise could pose additional storm surge or tsunami risk to heavily populated low-lying coastal regions such as lower Egypt, Bangladesh and many Pacific islands, along with the loss of most freshwater resources in the latter case. At higher latitudes, global warming may induce profound effects upon the human use and occupancy of land underlain by permafrost. Regardless of the causes, the impacts of anticipated changes in extreme weather hazards as a result of global climatic change, and their implications for human activity, demand the attention of geomorphologists.

Observational framework: natural hazards paradigms

Early academic research into natural hazards was characterized by an emphasis on human response to natural events. American geographer Gilbert White (1974) proposed the following research paradigm:

1 estimate the extent of human occupancy in areas subject to natural hazards;
2 determine the range of possible adjustment by social groups to those extreme events;
3 examine how people perceive the extreme events and resultant hazards;
4 examine the process of choosing damage-reducing adjustments;
5 estimate the effects of varying public policy upon that choice process.

This view emphasizes human response to specific catastrophic events, focusing on only extreme events, and implying that rational decisions are made based on cultural perceptions. Subsequent studies have increased the importance of risk assessment and the vulnerability of a population based upon the probability of an event. Although these views appear to reduce the role of the geomorphologist, the evaluation of a site with respect to specific risk lies at the very heart of hazard research.

Burton *et al.* (1978) suggest ranking the significance of potential hazards by evaluating the physical parameters of an event in terms that are obvious to geomorphologists:

1 magnitude: high to low
2 frequency: often to rare
3 duration: long to short
4 areal extent: widespread to limited
5 speed of onset: rapid to slow
6 spatial dispersion: diffuse to concentrated
7 temporal interval: regular to random.

Although qualitative, this view recognizes that events can range from intensive (such as a storm surge briefly affecting a stretch of coastline) to pervasive (such as the erosional effects of global sea-level rise).

Causal linkages are inherent in the notion of *geomorphic* hazards, where an extreme event may initiate other exceptional events of another type. Thus geomorphic hazards that are associated with landform response may be 'triggered' by climatic, hydrological, geophysical or man-induced events. Landslides may be causally linked with earthquakes, volcanic eruptions, heavy precipitation or construction activity. Pervasive linkages can result from land use change within a watershed affecting the magnitude and frequency of discharge within the stream.

Planners and developers often focus on a particular site or region, where mapping of hazard areas evaluates the potential risks for all potential hazards in such locations. Most physical scientists shun this 'hazardousness of place' concept, not wanting to venture beyond their own areas of expertise. Many social scientists characterize the actual hazard event only by its immediate physical effects, concentrating only on the societal response.

Geomorphologists recognize that the high-magnitude, low-frequency catastrophic events (large earthquakes, hurricanes) capture the attention due to the immediacy of large casualty and financial losses, but that events of moderate frequency (landslides, floods) often do as much or more collective damage. Geomorphic hazards tend to be more at the pervasive end of the hazards continuum, have slower speed of onset, longer duration, more widespread areal extent, more diffuse spatial dispersion, and more regular temporal interval. Exceptions such as slope failure exist, but in general landform change occurs over the long term at slow rates. Nevertheless geomorphologists should adopt a hazard paradigm in an effort to promote compatibility within this area of complex, and essential, interdisciplinary research.

Geomorphic hazards research

Gares *et al.* (1994) use the paradigm suggested by Burton *et al.* (1978) to discuss the role of geomorphic hazards research with respect to specific hazards. They illustrate the great variety of processes involved and suggest geomorphic evaluation in terms of the following aspects:

1 the dynamics of the physical process;
2 the prediction of the rate or occurrence;
3 the determination of the spatial and temporal characteristics;
4 an understanding of people's perception of the impact of the occurrence;
5 knowledge of how the physical aspects can be used to formulate adjustments to the event.

Geomorphologists vary widely in their definition of geomorphic hazards. Gares *et al.* (1994) limit their inclusion only to those process actions that gradually shape landforms, not agents of catastrophic change that arise from the consequences of geophysical, hydrological or atmospheric hazards, although many of these hazards result in geomorphic events. Wolman and Miller (1960) recognized that low-frequency, high-magnitude events often produce spectacular damage and geomorphic change, but events of moderate magnitude often do as much work (damage, change) cumulatively over the long run. As an encyclopedic entry, our definition will be necessarily inclusive of all agents of surficial change, pervasive to episodic.

Pervasive processes, such as soil erosion, are minimal in natural environments, but accelerate greatly with human disturbance such as forest clearing or agricultural tilling. Perception of soil erosion as a hazard involves farmers who lose crops to sheet wash and gullies, water managers and engineers who suffer siltation in reservoirs or canals, fishermen whose catch is reduced by silt and turbidity, and all who suffer from reduced crop and water yields. Soil erosion rarely results in direct loss of life, but it has a widespread distribution, high remediation costs, and long-term effects on water and food production. Despite more than seventy years of soil erosion mitigation research, countless thousands suffer malnutrition due to lost soil productivity.

Numerous geomorphic processes have causal links to volcanic eruptions. Geophysicists and volcanologists monitor eruptive precursors, prior eruptive history and distrubution of past eruptive products to assist disaster managers with warnings about the type and magnitude of imminent risk. Geomorphologists contribute to post-eruption mitigation efforts, as impacts such as pyroclastic flows and ash fall frequently result in unstable slopes, lahars clog stream channels, and overall sediment yield is greatly increased. Siltation and debris loading of streams radiating out from affected areas result in reduced channel capacity, increasing the frequency of overbank flooding, causing flow deflections and bank erosion. Applications of geomorphology include erosion control and engineering impact analysis, along with mitigation and recovery planning. Since the 1980s geomorphologists have made significant contributions following the eruptions of Mt. St Helens, USA, Mt. Pinatubo, Philippines and Nevada del Ruiz, Colombia.

Heavy rainfall can saturate soils causing rapid debris flows and mudflows. In 1938, such events in Japan were triggered by typhoon rains, resulting in the loss of more than 130,000 homes and over 2,000 lives. The magnitude of this loss prompted government attention focused on landslide control, and similar rains in 1976 affected less than 2,000 homes and cost 125 lives. Similar reductions have occurred with other catastrophic events. The horrific death toll experienced in Bangladesh due to storm surge and flooding in 1970 was reduced by thirty to fifty times during a cyclonic storm surge of similar magnitude in 1985 because a satellite-based early warning system prompted the evacuation of island and coastal dwellers. These two examples come from opposite ends of Asia's economic spectrum. Rosenfeld (1994a) points out that economically developed countries often suffer the greatest economic losses, while their lesser developed counterparts endure the highest loss of life. In developing nations, there is often a conscious decision to allocate resources toward economic development, at the risk of underfunding disaster mitigation, often with the effect of greater loss of both infrastructure and human lives.

In some instances, disaster mitigation strategies and international relief efforts may actually be partially responsible for rising losses. In many developed countries, state-backed hazard insurance programmes are designed to encourage the use of hazard zoning and the implementation of damage-resistant building codes to reduce the demand for structural control measures. However this may have actually encouraged the development of hazard-prone areas through the combined effect of lower land costs and cheap indemnity. Thus, the 'insured' transfers the risk to the 'insurer' and may dismiss the concern for loss prevention measures.

Geomorphologists have the opportunity to demonstrate the nature of geomorphic hazards (Figure 66), map landform or surface material conditions that have hazard potential, and recognize the effects of human modification of natural conditions which could result in increased hazard potential. As scientists, we are reluctant to translate this knowledge into arguments for the adoption of specific mitigation or management strategies, and thus are less than proficient at applying our information base. Most land use managers, planners, developers and government decision-makers rely on 'on the job' training to develop expertise in the interpretation of technical information for risk assessment and disaster reduction. Often this is prompted only in response to significant losses.

Automated monitoring networks and advanced computer modelling techniques are giving us new tools to test alternative hazard mitigation strategies. Geomorphologists must be willing to embrace new technologies which will permit them to exercise their specialized talents globally, interface more readily with professionals outside the Earth science community (for example social scientists and engineers), and associate more closely with monitoring networks and scientific unions to ensure that major events are anticipated by identifying their physical precursors. Natural hazards research is obviously an interdisciplinary field involving a range of physical scientists with social scientists assessing the human dimension of the problems. Given the limited observed record of hazard events in most regions, a geomorphological approach, where the areal extent, and perhaps the frequency, of events can be determined from the landscape, is essential. The geomorphological approach may also encourage the 'nature knows best' path to designing hazard mitigation strategies in balance with the dynamics of processes within the region. In the final analysis, occupance of hazard-prone areas is both physically and economically self-regulating. It is the function of science, as a servant of society, to identify those limitations and point the way toward minimizing the disastrous consequences of 'learning our lessons' in nature's way.

References

Berz, G. (1993) The insurance industry and IDNDR: common interests and tasks, *IDNDR Newsletter 15*, Observatorio Vesuviano, 8–11.

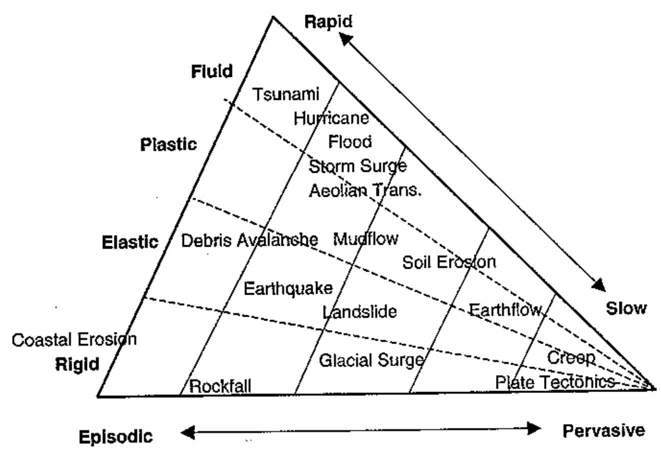


Figure 66 Geomorphic hazards include pervasive processes, which may be imperceptible to individuals, to episodic events which may have frequencies of occurrence below thresholds deemed significant by planners, but which may have significant magnitude. These processes involve the full spectrum of stress/strain modulus, and involve virtually every speciality within the discipline

Bruce, J.P. (1993) Natural disasters and global change, *IDNDR Newsletter 15*, Observatorio Vesuviano, 3–8.

Burton, I., Kates, R.W. and White, G.F. (1978) *The Environment as Hazard*, Oxford: Oxford University Press.

Carrara, A. and Guzzetti, F. (eds) (1995) *Geographical Information Systems in Assessing Natural Hazards*, Dordrecht: Kluwer.

Gares, P.A., Sherman, D.J. and Nordstrom, K.F. (1994) Geomorphology and Natural Hazards, *Geomorphology* 10, 1–18.

Rosenfeld, C.L. (1994a) Flood hazard reduction: GIS maps survival strategies in Bangladesh, *Geographical Information Systems* 2(3), 29–39.

——(1994b) The geomorphological dimensions of natural disasters, *Geomorphology* 10, 27–36.

Simpkin, T., Seibert, T., McClelland, L., Bridge, D., Newhall, C. and Latter, J. (1981) *Volcanoes of the World*, Stroudsburg, PA: Hutchinson Ross.

White, G.F. (1974) Natural hazards research: concepts, methods and policy implications, in G.F White (ed.) *Natural Hazards: Local, National, Global*, 3–16, Oxford: Oxford University Press.

Wolman, M.G. and Miller, J.P (1960) Magnitude and frequency of forces in geomorphic processes, *Journal of Geology* 68, 54–74.

CHARLES L. ROSENFELD

GEOMORPHOLOGICAL MAPPING

Geomorphological mapping encompasses one of a group of techniques under the general category of TERRAIN EVALUATION employed to record systematically the shape (or morphology), landforms, landscape-forming processes and materials that constitute the surface of the Earth. Lee (2001) identifies three forms of geomorphological map:

1. Regional surveys of terrain conditions, either for land use planning or in baseline studies for environmental impact assessment (e.g. the 1:25,000 scale maps of Torbay, Doornkamp (1988)).
2. General assessments of resources or geohazards at scales between 1:50,000 and 1:10,000 (e.g. Bahrain Surface Materials Resources Survey, Doornkamp *et al.* (1980); ground problems in the Suez City area, Egypt, Jones (2001)).
3. Specific-purpose large-scale surveys to delineate and characterize particular landforms (e.g. the 1:500 scale investigations around the Channel Tunnel portal, Folkestone, Griffiths *et al.* (1995)).

The initial stage of geomorphological mapping involves factually recording ground shape through a process of morphological mapping. This requires the production of a map on which the land surface is subdivided into planar facets separated by gradual changes or sharp breaks in slope. On the map the changes and breaks in slope are identified as either concave or convex in nature and recorded using decorated lines, a system first established by Savigear (1965). Arrows with a numeric value in degrees indicate the slope angle and downslope direction of the planar facets. Once the morphology has been recorded a geomorphological interpretation is undertaken whereby details of the contemporary and relict landforms and geomorphological processes are added to the map. In addition, data on the nature of materials and hydrology of the area are noted. Geomorphological interpretation can allow a suite of derivative maps to be produced, e.g. resource maps and landscape genesis maps. Standard symbols to be used on all these maps are contained in Cooke and Doornkamp (1990), although, Demek and Embleton (1978) provide a more comprehensive collection of symbols that allow subtle differences in the landscape to be highlighted. However, in many situations the geomorphological maps are produced as unique products with a bespoke legend.

The techniques used to compile the data involve both field survey and, where possible, examination of remote sensing information. The main form of remote sensing analysis has traditionally been through the interpretation of vertical pairs of aerial photographs viewed stereoscopically. An initial preliminary morphological map and geomorphological interpretation is produced using aerial photographs but this should normally be subject to 'ground-truth' mapping in the field. With the advent of higher resolution satellite images this preliminary mapping stage increasingly is being carried out using data from the array of new satellite-based scanners.

A two-person team normally undertakes the field mapping. The main requirement for the production of effective geomorphological maps is an accurate base map at a suitable scale. The base map may be a standard survey map depicting man-made and natural features including ground topography, or a spatially corrected ortho-photo. The field data should be compiled directly on the base map. Spatial data and slope information can be obtained through a simple tape, compass and clinometer survey, using more sophisticated land survey techniques, use of global positioning systems, or a suitable combination of these methods. The geomorphological, materials and hydrological data are noted on maps and recorded in field notebooks where appropriate.

Whilst geomorphological mapping has been used for general landscape investigations, it has been employed most successfully by applied geomorphologists, particularly for engineering studies. Brunsden *et al.* 1975, articulated the aims of geomorphological mapping for highway engineering:

1 Identification of the general terrain characteristics of the route corridor, including suggestion of alternative routes and location of hazards.
2 Defining the 'situation' of the route corridor, for example identifying influences from beyond the boundary of the corridor.
3 Provision of a synopsis of geomorphological development of the site, including location of materials for use in construction and location of processes affecting safety during and after construction.
4 Definition of specific hazards, e.g. landsliding, flooding, etc.
5 Description of drainage characteristics, location and pattern of surface and subsurface drainage, nature of drainage measures required.
6 Slope classification, according to steepness, genesis and stability.
7 Characterization of nature and extent of weathering, also susceptibility to mining subsidence and erosion.
8 Definition of geomorphological units, to act as a framework for a borehole sampling plan and to extend the derived data away from the sample points.

Although these aims were developed specifically for highway projects they represent an appropriate checklist for all geomorphological mapping programmes undertaken for civil engineering projects.

References

Brunsden, D., Doornkamp, J.C., Fookes, P.G., Jones, D.K.C. and Kelly, J.M.N. (1975) Large scale geomorphological mapping and highway engineering design, *Quarterly Journal of Engineering Geology* 8, 227–253.

Cooke, R.U. and Doornkamp, J.C. (1990) *Geomorphology in Environmental Management*, 2nd edition, Oxford: Oxford University Press.

Demek, J. and Embleton, C. (eds) (1978) *Guide to Medium-scale Geomorphological Mapping*, Stuttgart: International Geographical Union.

Doornkamp, J.C. (ed.) (1988) *Planning and Development: Applied Earth Science Background, Torbay*, Nottingham: MI Press.

Doornkamp, J.C., Brunsden, D., Jones, D.K.C. and Cooke, R.U. (1980) *Geology, Geomorphology and Pedology of Bahrain*, Norwich: GeoBooks.

Griffiths, J.S., Brunsden, D., Lee, E.M. and Jones, D.K.C. (1995) Geomorphological investigation for the Channel Tunnel and Portal, *Geography Journal*, 161, 257–284.

Jones, D.K.C. (2001) Ground conditions and hazards: Suez City development, Egypt, in J.S. Griffiths (ed.) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication 18, 159–170.

Lee, E.M. (2001) Geomorphological mapping, in J.S. Griffiths (ed.) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication 18, 53–56.

Savigear, R.A.G. (1965) A technique of morphological mapping, *Annals of the Association of American Geographers* 53, 514–538.

Further reading

Fookes, P.G. (1997) Geology for engineers: the geological model, prediction and performance, *Quarterly Journal of Engineering Geology* 30, 290–424.

JAMES S. GRIFFITHS

GEOMORPHOLOGY

Definition and scope

Geomorphology is the area of study leading to an understanding of and appreciation for landforms and landscapes, including those on continents and islands, those beneath oceans, lakes, rivers, glaciers and other water bodies, as well as those on the terrestrial planets and moons of our Solar System. Contemporary geomorphologic investigations are most commonly conducted within a scientific framework (see Rhoads and Thorn 1996) although academic, applied or engineering interests may motivate them. A broad range of alternative research methodologies have been employed by geomorphologists, and past attempts to impose a systematic structure on the discipline have yielded stifling tendencies and overt resistance. Geomorphologists frequently profess to innate aesthetic appreciation for the complex diversity of Earth-surface forms, and, in this

regard, a fitting definition of geomorphology is simply 'the science of scenery' (Fairbridge 1968).

Past and present concerns have focused on the description and classification of landforms (including their geometric shape, topologic attributes and internal structure), on the dynamical processes characterizing their evolution and existence, and on their relationship to and association with other forms and processes (geomorphic, hydro-climatic, tectonic, biotic, anthropogenic, extraterrestrial, or otherwise). Geomorphology is an empirical science that attempts to formulate answers to the following fundamental questions. What makes one landform distinct from another? How are different landforms associated? How did a particular landform or complex landscape evolve? How might it evolve in the future? What are the ramifications for humans and human society?

Modern geomorphology is currently subdivided and practised along the lines of specialized domains. *Fluvial geomorphology*, for example, is concerned with flowing water (primarily in the form of rivers, streams and channels) and the work it accomplishes during its journey through the terrestrial phase of the hydrologic cycle. A very broad spectrum of interests are subsumed within fluvial geomorphology, ranging from the influence of turbulence on the entrainment, transport and deposition of sediment particles at the finest scale, to the mechanics of MEANDERING, POINT BAR formation and FLOODPLAIN development at middle scales, to the nature and character of DRAINAGE BASIN evolution at the coarsest scales. Within the other substantive areas of geomorphology are: *hillslope geomorphologists*, who boast expertise on the geotechnical properties of soil and rock, the mechanics of LANDSLIDES, and the movement of water within the ground; *tectonic geomorphologists*, who study neotectonic (see NEOTECTONICS) stress fields, continental-scale sedimentary basins and active/passive margin landscapes; *glacial and periglacial geomorphologists*, who are interested in alpine and continental glaciers, PERMAFROST and other cold-climate forms or processes that involve ice, snow and frost; *karst geomorphologists*, who deal with soluble rocks (e.g. limestone) and chemical processes of DISSOLUTION that lead to landforms such as gorges, caverns and underground streams; *coastal geomorphologists*, who study nearshore, lacustrine and marine systems where oscillatory, rather than unidirectional, flow processes dominate; and *aeolian geomorphologists*, who study the transport of sand and dust by wind, mostly in desert or semi-arid environments, but also along beaches, over agricultural fields and on the moon and Mars. Other subspecialities include: soils geomorphology, biogeomorphology (zoogeomorphology), climatic geomorphology, tropical geomorphology, desert geomorphology, mountain geomorphology, extraterrestrial (planetary) geomorphology, remote-sensing geomorphology, experimental geomorphology, environmental geomorphology, forest geomorphology, applied geomorphology, engineering geomorphology and anthropogeomorphology.

Major themes and concepts

Landforms are dynamic entities that evolve through time as a consequence of characteristic suites of processes acting upon Earth-surface materials. Geomorphologists are concerned with documenting and unravelling the mysteries of this process-form interaction. Relevant knowledge includes not only the manner and direction of landform evolution (progressive or cyclic, slow or rapid), but also the processes that dominate or direct the evolution (type, intensity) as well as the mutual adjustments and feedbacks that occur between the forms and processes as energy and matter are cycled through the landscape. To better understand these complex interrelationships, geomorphologists have proposed various conceptual themes or templates to aid in organizing their thinking. Among these are:

1 *Endogenic–exogenic forces* Geomorphic systems are governed by dynamic controls that may be internally produced (endogenic) or externally imposed (exogenic) upon the system. Tectonic, volcanic and isostatic activities are manifestations of endogenic forces within Earth, whereas rainfall and meteorite showers are exogenic forces. The spatial and temporal scales of the geomorphic system influence the types of endogenic–exogenic forces that are relevant. The control–volume, force–balance approach in fluid mechanics is an analogue to this concept.
2 *Destructive–constructive action* Some geomorphic processes create landforms (e.g. volcanic cones, meteorite impact craters, termite mounds) whereas other processes (e.g. chemical weathering, rainwash, human activity) destroy landforms or cause widespread denudation. More typically, most geomorphic

processes both create and destroy landforms simultaneously. For example, flowing water in a river will erode the outer bank of a meander bend while depositing sediment on the inner bank in the form of a point bar. Similarly, a glacier can erode, excavate and sculpt the surface upon which it moves while also depositing sediment in the form of eskers and moraines.

3 *Erosional–depositional forms* Some landforms are sculpted by erosion of pre-existing materials (e.g. bedrock canyons, roches moutonnées) whereas others are built via deposition of new material on existing substrate (e.g. deltas, lava flows). Yet others are hybrid features formed through both erosion and deposition locally (e.g. impact craters) or maintained by an intricate balance between erosion and deposition at different positions on the same form (e.g. migrating sand dunes).

4 *Stress–strength relationships* Most geomorphic processes induce landscape change by stressing the system, as with flowing fluids, chemical reactions, tectonic motion or the prolonged action of gravity. The materials upon which these processes act have the ability to resist change because of inherent properties that provide strength (e.g. mineralogy, cohesion, structure, relative placement). Geomorphologists have generally devoted more effort to measuring processes than to investigating how material strength is a complementary and counterbalancing factor (exceptions are many, and they include the efforts of the Japanese school to elucidate systematically the nature of rock control on geomorphic evolution, of the many coastal geomorphologists interested in rocky coasts, of several geomorphological geologists concerned with relict landforms and ancient landscapes, and various engineering geomorphologists who study slope failures).

5 *Polygenesis and inheritance* Landscapes consist of landform assemblages that are rarely simple. Complex suites of polygenetic forms may coexist in the same location if, for example, multiple processes are active contemporaneously or when particularly resistant relict forms are inherited from prior eras. The latter are progressively modified by contemporary processes that also create new forms to produce palimpsest landscapes.

The integrated sum of exogenic–endogenic forces and destructive–constructive actions working in concert to create erosional–depositional forms according to dominant stress–strength relationships will dictate whether landscape RELIEF will be enhanced or reduced over a given time interval. At one end of the spectrum are the steep mountain and valley systems of the globe (e.g. Himalayas), and at the other are the extensive abyssal plains of the deep ocean as well as the PENEPLAIN of William Morris Davis. Geomorphologists have identified several additional themes and concepts that serve to strengthen the theoretical foundation of their science. These include scale, causality, equilibrium, equifinality, thresholds, magnitude–frequency, landscape memory and relaxation. Readers should consult the references and other entries in this encyclopedia for detailed discussions.

Early historical development

The subject matter of geomorphology has occupied human thinking for thousands of years, and early writings on landforms can be traced to the time of the ancient Greek, Roman, Arab and Chinese philosophers. Aristotle (384–322 BC) and Strabo (54 BC–AD 25), for example, had keen insights into the origin of springs, the work of rivers and the importance of earthquakes and volcanoes. Nevertheless, the history of geomorphology (see Chorley *et al.* 1964; Tinkler 1985, 1989) is typically traced back only as far as the European Renaissance because few written documents about geomorphic knowledge remain from the period prior to the sixteenth century. During the Renaissance, most studies of Earth were conducted from a naturalist, philosophical perspective because specialized academic disciplines had not evolved and scientific methods were not widely known. Leonardo da Vinci, Bernard Palissy, Nathanael Carpenter, Bernhard Varenius, Thomas Burnet and Nicolaus Steno are among the key figures from this period, and unwittingly they began to lay the foundation for the science of geomorphology. Unfortunately, this was also a time when the Church exerted powerful control over academic thinking, and the predominant objective of learned men was to reconcile their day-to-day observations of natural processes with strict religious orthodoxy and bibliolatry. The biblical scholar, James Ussher, Archbishop of Armagh, decreed that Earth was created on Sunday, 23 October 4004 BC and that the Flood of the Old Testament began in 2349 BC, and in so doing, he may well have imposed the most stifling proclamation on the developmental history of the Earth sciences. All evolutionary processes, by definition, had now to be contemplated within the constraints of a 6,000-year Earth history, and to think otherwise was heresy. Unsurprisingly, the dominant interpretation of Earth-surface processes invariably involved catastrophes, cataclysms and disasters such as global deluges and seismic convulsions.

The period following the Renaissance and into the early nineteenth century was one of scepticism, controversy and debate. It was also one that witnessed several changes that bear directly on the development of geomorphology as an academic discipline. The first was the evolution of specialized areas of study such as biology, physics, astronomy, mathematics, hydraulics and geology, and this set the stage for various subdisciplines, such as petrology, mineralogy, paleontology, stratigraphy and geomorphology, to be spawned. Second was a slow transformation in academic discourse away from the unassailable validity of belief systems and authoritarianism toward a standard of proof based on empiricism and observable evidence. Third was the development of increasingly sophisticated instrumentation and measurement technologies and protocols. Fourth was the enhanced mobility of people and information, thereby facilitating greater exposure to new and interesting environments and ideas. And fifth was the gradual acceptance of gradualism (see UNIFORMITARIANISM) in favour of CATASTROPHISM. Two dominant factions emerged during this period. The Neptunists (or Wernerians) followed the ideas of a German mineralogist, Abraham Gottleb Werner, who contended that rocks on Earth originated from mechanical and chemical processes in a universal ocean. The Plutonists (or Vulcanists) stressed the importance of intrusive and extrusive volcanic processes in rock formation. Key figures during this period include members of the 'French School', such as Jean Étienne Guettard, Nicolas Desmarest and Jean-Baptiste Lamarck, as well as the Swiss geologist, Horace Benedict de Saussure.

James Hutton, credited by some as the founder of modern geomorphology, was a Plutonist who argued vehemently for the importance of gradual subaerial denudation across millennia. His uniformitarian ideas, expressed in well-known phrases such as 'the present is the key to the past' and 'no vestige of a beginning, no prospect of an end', were revolutionary because they shifted the focus of attention away from catastrophic events of 'creation' toward continuous, everyday agents of erosion. Unfortunately, Hutton's teachings were not warmly received by the conservative cognoscenti of that time. After Hutton's death, his friend and colleague, John Playfair, published a book that explained and expanded Hutton's writings, and by the beginning of the nineteenth century, a slow conversion to gradualism was taking hold. Cyclic and timeless theories of landscape evolution were coming into vogue. Three schools of thought regarding landform evolution emerged. DILUVIALISM represented a transformed extension of the catastrophist lineage, and diluvialists such as Reverend William Buckland and Reverend Adam Sedgwick believed that huge floods carved many surface features. Structuralists, such as Henry Thomas de la Beche and John Phillips contended that structural controls were paramount to understanding landscape genesis (see STRUCTURAL LANDFORM), while also acknowledging that both catastrophic and gradual processes could yield substantial erosion. Fluvialists, in contrast, argued for the dominance of rivers and streams in wearing away the landscape through slow, but continuous action.

A chief proponent of fluvialism and UNIFORMITARIANISM was Sir Charles Lyell, whose *Principles of Geology* went into twelve editions after original publication in 1830. Lyell based his arguments on careful observations and measurements, and effectively attacked the notions of theological reconciliation, catastrophism and diluvialism. His writings on uniformitarianism incorporated four distinct notions: (1) uniformity of law (the laws of nature are immutable); (2) uniformity of process (processes operative today were also operative in the past, and exotic causes need only be invoked unusually); (3) uniformity of rate (gradualism); and (4) uniformity of state (change is endlessly cyclical and directionless). The publication of Lyell's *Principles* engendered considerable debate, and the period through to the middle of the nineteenth century witnessed both conflict and compromise regarding the importance of fluvial action, pluvial denudation, marine dissection, iceberg drift and glaciation (see GLACIAL THEORY) as agents of erosion. Indeed, even Lyell began to expound the virtues of marine dissection above fluvial degradation. In part, this was due to the existence of various unexplainable observations such as major unconformities in the stratigraphic record and huge ERRATICS in unexpected places. The powerful action of the sea presented an expedient solution because submarine processes could not be observed or measured directly, and theorization could proceed unbridled. Nevertheless, most of these seemingly contradictory theories incorporated at least some common elements and themes and, invariably, they were cast within a framework of uniformitarianism rather than catastrophism.

By the mid-1870s, some consensus was beginning to emerge about the multifaceted and complex nature of landscape evolution. The marine planation theory of Sir Andrew Crombie Ramsay, for example, proposed that the action of waves and currents in the ocean was not to dissect the sea bottom, but to level off bathymetric protuberances thereby producing marine plains. Upon emergence through tectonic activity, subaerial forces become active and fluvial erosion proceeds to carve out valleys and denude landscapes. Support for this theory came from the many accordant summit heights in the highlands of Wales and England, as well as from the marine abrasion studies of Baron Ferdinand von Richthofen in China. Concurrently, the glacial theories (see GLACIAL THEORY) of Ignace Venetz, Jean de Charpentier and Louis Agassiz were receiving widespread acceptance decades after their introduction, albeit with climatic and glaciofluvial amendments. This was a significant development in geomorphology because environmental dynamism (see DYNAMIC GEOMORPHOLOGY) was implicit to these theories. Gradualist and neo-catastrophist perspectives could both be accommodated under this new framework because uniformity of process (the nature of past and present processes are the same) did not necessarily imply that the intensities and rates of process action could not vary.

At the conclusion of the nineteenth century, geomorphology was poised to begin its emergence as a modern scientific discipline. The word 'geomorphology' had already been coined in the mid-1800s (Tinkler 1985: 4), and several textbooks on exclusively geomorphic matters had been written. As an area of academic study, geomorphology was experiencing legitimate interest under the guise of 'physiography' or 'physiographical geology'. Centres of expertise were arising in many different countries within and outside Europe, all with subtly different identities and separate agendas. British geomorphologists, for example, spent considerable effort on compiling complex denudation chronologies linked to marine processes and periods of tectonic stability/instability and sea-level fluctuation. German geomorphologists (e.g. Hettner, A. Penck, Walther) became interested in the influence of climate as a consequence of conducting research in the Alps as well as in the subhumid tropics. The North American school, in contrast, was dominated by fluvialism buoyed by indisputable evidence derived from the great explorations of the largely unvegetated, semi-arid West. John Wesley Powell's trips into the Grand Canyon and his reports on the Colorado Plateau and Uinta Mountains provided powerful testimony to the efficacy of rivers to erode landscapes. Grove Karl Gilbert's studies on the mechanics of fluvial erosion, sediment transport and turbulence are exemplars of the elegant application of the scientific method. He also investigated the origin of pediments and lateral planation, and in recognition of his many contributions, Gilbert is often identified as the first truly process-oriented American geomorphologist. Indeed, it is largely due to the efforts of Powell, Gilbert, Dana and Dutton and various other United States Geological Survey employees that the North America school became the dominant force in the development of geomorphology at the turn of the century.

Twentieth-century developments

Geomorphology in the twentieth century experienced rapid evolution and growth, and six overlapping phases of development can be identified. These are little more than crude caricatures, and the reader is referred to Chorley *et al.* (1973) and Beckinsale and Chorley (1991) for detailed discussions of the key figures and their substantive contributions. The *historical* phase, roughly from 1890–1930, was dominated by William Morris Davis and his many disciples. Davis's deductively derived model, 'The Geographical Cycle', envisioned serial evolution of landscapes beginning with rapid tectonic uplift followed by progressive denudation in characteristically distinct stages of 'youth', 'maturity' and 'old age'. It served as the genetic template upon which reconstructive narratives of landscape evolution were hung, with relatively little concern for the mechanical and chemical processes responsible for erosion and deposition. Nevertheless, these denudation chronologies spawned keen interest in tectonic geomorphology as well as an appreciation for the importance of unravelling the historical sequence of steps that ultimately manifest themselves as a contemporary landscape.

The *regionalist* phase (1920–1950) was characterized by detailed and thorough investigations of regional landscapes, both in the conventional mid-latitudes of North America and Europe as well as in globally remote areas (e.g. tropics, deserts, high latitudes). Increasingly, these regionally based studies yielded data about landforms and landform assemblages that could not be easily explained within the framework of Davis's geographical cycle, especially his contentions about the 'normalcy' of the humid, mid-latitudes. Although Davis found support within Britain and France, many European schools remained unconvinced by Davis's teachings. Walther Penck, for example, proposed an alternative model of landscape evolution that highlighted the importance of the relative rates of uplift and denudation in controlling landform geometry. Another German geomorphologist, J. Büdel, stressed the dominance of climatic controls and proposed the concept of etchplanation (see ETCHING, ETCHPLAIN AND ETCHPLANATION) and MORPHOGENETIC REGIONS. Climatic geomorphology was also practised by Louis Peltier in North America and it was later championed by J. Tricart and A. Caillieux in France. In this way, Davis's unifying ideas gradually fell into disfavour, and geomorphology became an empirically driven scientific confederacy of polyglot regionalist schools.

The *quantitative* phase (1940–1970) reflected a broader trend within many of the Earth sciences toward enhanced use of sophisticated technologies (often derived from the war effort) to measure, describe and analyse the surface features of Earth. R.E. Horton's publications on stream networks and drainage basin processes are classically identified as the precursor to this quantitative movement, but the foundational works of Bagnold, Gilbert, Hjulstrom, Leighly, Rubey and Shields, among many others, are rightfully acknowledged. These early 'quantifiers' were concerned to understand landforms and geomorphic processes in deterministic or probabilistic, but testable, ways rather than on the basis of deductively derived heuristic models that ultimately yielded little predictive power. Logical positivism was the dominant philosophy and reductionism was the overriding methodological approach. As a consequence, geomorphology became increasingly fragmented and specialized, with fewer and fewer connections between the sub-specializations as well as pronounced distancing from its mother disciplines of geography and geology. Fortunately, connections to other allied disciplines such as fluid mechanics, engineering hydrology, statistics, thermodynamics, meteorology, pedology and agricultural physics were being cultivated, and these provided a theoretical and conceptual richness upon which geomorphologists could draw, if so inclined.

The *systems* phase (1960–1980) in the development of geomorphology was inaugurated by the introduction of general systems theory into the conceptual toolkit of geomorphology by Richard J. Chorley, which was a logical outgrowth of the quantitative phase. The quantification of prior decades was basically of two genres: (a) statistical 'black-box' description (e.g. Horton's Law of Stream Numbers); and (b) detailed measurement and interpretation of dynamical processes (e.g. Strahler 1952). The former proved unrewarding in terms of providing insight into geomorphic behaviour, whereas the latter were typically conducted at a scale that was too small to be relevant to landscape evolution. The systems approach alleviated the 'black-box' quandary by describing geomorphic behaviour in terms of energy and mass flows, equilibrium tendencies, relaxation times and thresholds (see THRESHOLD, GEOMORPHIC) of response. A large number of concepts, such as ALLOMETRY, entropy and ergodicity (see ERGODIC HYPOTHESIS), were borrowed from other disciplines as theoretical templates. These were applied to a broad range of geomorphic systems with varying degrees of success, but the large number of journal articles and textbooks containing box-and-arrow plots attests to the popularity of this approach during the systems phase. Unfortunately, there was an irresistible tendency to equate system behaviour with geomorphic process, much to the detriment of dynamical process investigations.

Since about the 1980s, geomorphology has entered a phase of increasing *reconciliation* and *unification* that signals its arrival as a mature modern science. Introspective debates about catastrophic versus uniformitarian ideas, quantitative-deterministic/stochastic versus qualitative- historical methodologies, and geographical versus geological disciplinary roots are taking place not for purposes of disciplinary leadership or hegemonic posturing, but rather in consequence of the pragmatic need for geomorphology to assert an identity distinct from that of other Earth sciences (e.g. geology, geography, sedimentology, stratigraphy, paleontology) as well as to understand the complex spectrum of conceptual ideas upon which geomorphology is founded (e.g. Rhoads and Thorn 1996). Many extremist ideas of prior eras have been reintroduced into the literature as softened compromises (e.g. NEOCATASTROPHISM, neo-historicism, neo-regionalism) to provide balance to the uniformitarian-style fluvialism that dominated the quantitative and system phases. Invariably, these conceptual ideas were discussed in the context of factual evidence and with a view toward generating insight into unusual geomorphic features or terrain that belie conventional explanation (e.g. Baker 1981). The modern-day geomorphologist has a deep appreciation for the importance of slowly acting processes in concert with large-magnitude, low-frequency events in leaving imprints on the landscape, for the utility of detailed process-mechanical studies as well as historical reconstructions of landform assemblages in unravelling the complexities of the present-day surface, for the interconnectivity between the various sub-specializations of geomorphology and allied Earth and engineering sciences, and for the complementarities among twenty-first century technological capacities when combined with a field geomorphologist's keen sense of the lie of the land.

Future directions

Geomorphology in the twenty-first century will continue to mature as a science and assert its importance among the Earth-science disciplines. The issue of scale will remain a dominant topic of investigation and discourse, and it will be richly informed by expanding concerns about tectonic and structural controls on geomorphic systems over long time periods (i.e. megageomorphology), the evolution of lunar and Martian surfaces (i.e. planetary geomorphology), the intricate linkages between geomorphic and biogeochemical systems, and the hierarchically nested versus scale-invariant nature of geomorphic systems. The term 'neogeomorphology' was recently coined (Haff 2002) to suggest that a new or modern form of geomorphology may be evolving – one which, of necessity, takes into account the sobering fact that humans now displace more soil and rock per year than rivers, glaciers and wind combined (Hooke 2000). The pace of anthropogenically driven landscape alteration, whether direct or indirect (e.g. via global warming), is likely only to increase in the future. And, because there are no analogues for such pronounced surface modification in the stratigraphic record, the relevance and utility of geomorphology (with its traditional focus on process–form interaction) to the planning and environmental management communities seems assured. Geomorphologists already play central roles in mandated environmental impact assessments involving construction, mining and forestry, and increasingly their expertise (in conjunction with biologists and botanists) is utilized in landscape reclamation, rehabilitation and restoration efforts involving streams, wetlands and coastal dunes.

In the quest for a deeper understanding of the past (retrodictive) and future (predictive) evolution of Earth's surface, geomorphologists are becoming increasingly reliant on sophisticated technologies. These include: new dating methods (e.g. cosmogenic radionuclides, optical- and thermo-luminescence, rock varnish, lichenometry) that yield the relative ages of landform elements and thereby unfold the historical sequence of events that produced the landscape; novel remote-sensing techniques (e.g. interferometric synthetic-aperature radar, lidar, ground-penetrating radar, time-domain reflectrometry) to measure and monitor a broad range of surface and subsurface attributes; advanced computational methods involving more powerful hardware, more efficient software codes, and more easily integrated and interoperable data platforms (e.g. Geographical Information Systems, Digital Elevation Models); and enhanced satellite coverage to provide synoptic information about inaccessible regions and across large distances with ever-increasing accuracy regarding absolute location and relative movement via the Global Positioning Systems. In addition, the means to communicate information and ideas virtually instantaneously to the entire community of geomorphologists has been greatly facilitated by the World Wide Web and by various national and international organizations such as the International Association of Geomorphologists (IAG) that maintain electronic bulletin boards and membership/address lists. For the first time in its long developmental history, geomorphology has the potential to become a truly global enterprise in terms of both coverage and participation.

References

Baker, V.R. (ed.) (1981) *Catastrophic Flooding: The Origin of the Channelled Scablands*, Stroudsburg, PA: Dowden, Hutchinson and Ross.
Beckinsale, R.P. and Chorley, R.J. (1991) *The History of the Study of Landforms or the Development of Geomorphology: Volume 3, Historical and Regional Geomorphology 1890–1950*, New York: Routledge.
Chorley, R.J., Beckinsale, R.P. and Dunn, A.J. (1973) *The History of the Study of Landforms or the Development of Geomorphology: Volume 2, The Life and Work of William Morris Davis*, London: Methuen.
——Dunn, A.J. and Beckinsale, R.P. (1964) *The History of the Study of Landforms or the Development of Geomorphology: Volume 1, Geomorphology Before Davis*, London: Methuen.
Fairbridge, R.W. (1968) *The Encyclopedia of Geomorphology*, New York: Reinhold.
Haff, P.K. (2002) Neogeomorphology, *EOS, Transactions of the American Geophysical Union* 83(29), 310.
Hooke, R. LeB. (2000) On the history of humans as geomorphic agents, *Geology* 28, 843–846.
Rhoads, B.L. and Thorn, C.E. (eds) (1996) *The Scientific Nature of Geomorphology*, Chichester: Wiley.
Strahler, A.N. (1952) Dynamic basis for geomorphology, *Geological Society of America Bulletin* 63, 923–938.
Tinkler, K.J. (1985) *A Short History of Geomorphology*, London: Croom Helm.
——(ed.) (1989) *History of Geomorphology, from Hutton to Hack*, London: Unwin Hyman.

Further reading

Leopold, L.B., Wolman, M.G. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: Freeman.
Ritter, D.F., Kochel, R.C. and Miller, J.R. (1995) *Process Geomorphology*, 3rd edition, Dubuque, IA: William C. Brown.
Scheidegger, A.E. (1970) *Theoretical Geomorphology*, Berlin: Springer-Verlag.
Schumm, S. (1991) *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge: Cambridge University Press.
Yatsu, E. (2002) *Fantasia in Geomorphology*, Tokyo: Sozosha.

BERNARD O. BAUER

GEOMORPHOMETRY

Dealing with quantitative analysis of the land surface, geomorphometry is a central theme in both theoretical and applied geomorphology (Pike and Dikau 1995). It is also diverse, dealing both with landforms and with the land surface as a one-sided rough surface, with vertical position a unique function of horizontal location. It could also be referred to as the combination of 'landform morphometry' and 'land surface morphometry'. It does not include surveying, photogrammetry and profiling, which provide the raw data for geomorphometry, but some knowledge of these is essential in considering error margins (Richards 1990: 36–41). Morphometry itself is a broader field, important not only in various aspects of Earth science, but also in engineering, biology and medicine. Each field of application has things to teach the others (Pike 2000), so long as the specifics of the original application are remembered. Geomorphometry inspired the idea of statistical FRACTALS, following difficulties in specifying 'how long is a coastline?'

Where individual landforms can be defined and distinguished from their surroundings, a series of MORPHOMETRIC PROPERTIES can be measured to provide a multivariate characterization of the landform. Such analysis is labelled 'specific geomorphometry'. This is distinct from 'general geomorphometry' of the land surface as in spectral or fractal analysis, or the study of surface derivatives and their interrelations (Evans 1980). General geomorphometry was extremely difficult before the introduction of computers: today it may involve the processing of very large DIGITAL ELEVATION MODELS (DEMs), and has many applications in digital terrain modelling (Pike 2000). Specific geomorphometry has a much longer history, starting with measurements of lunar craters and of coastal sinuosity in the nineteenth century.

The two aspects of geomorphology are not completely distinct, first because some specific landforms such as slopes or hillslopes and drainage networks are so widespread on Earth that their specific geomorphometry acquires a general importance: and second, because some techniques of general geomorphometry can be applied to specific landforms (Evans 1987). This permits analysis of variation within a landform (distributional analysis), not just generalization of its overall characteristics, and is more useful in the context of modelling.

General geomorphometry; surface derivatives

General geomorphometry starts with the altitude (elevation) of the surface – its height above sea level. This has major effects on climate and thus on surface processes. The frequency distribution of altitude (hypsometry) tells us quite a lot about the land surface. In the pre-computer era, this was summarized by its range (relief) and by the hypsometric integral – the relation of mean altitude above minimum, to this range. Relief varies with the size of area considered, and reaches several km (ridge to valley) in high mountain areas: the total range for the Earth is 8,852 + 11,033 m, i.e. 19.9 km. Hypsometric integral is around 0.5 for topography with sharp ridges and valleys, approaching but not reaching 1.0 for a plateau with few deep valleys, or 0.0 for a lowland with a few high hills. Evans (1972) suggested that instead of ranges and extremes, the use of standard statistical concepts – standard deviation and skewness – was both more economic and provided more stable statistics, influenced by the whole body of the distribution rather than by the extremes.

Ohmori (1993) found that hypsometric curves of mountainous areas such as Japan tend to be S-shaped or concave, giving integrals between 0.15 and 0.50. They can be simulated from empirically based relations between uplift, altitude, altitude dispersion and denudation rates. Hypsometric curves vary considerably with extent of area considered, and whether headwaters, large erosional basins or areas including depositional plains are analysed. Fuller understanding of landscape development is obtained by considering dimensional indices (mean and standard deviation of altitude) and not just dimensionless indices.

Gradient (slope angle) is the second local value of a surface that is very important in geomorphology and hydrology. It provides the stress to generate mass movements, and gives energy to surface flows. Engineers prefer percentages, i.e. 100 × (tangent of angle), but geomorphologists prefer to measure gradient in degrees. Mean gradient is of primary interest, but standard deviation and skewness of the point-by-point distribution of gradients tell us much about the regional topography. Fluvially dissected hill areas with slopes near some threshold tend to have low standard deviations of gradient, while glaciated mountains with cliffs, valley floors, terraced areas and often plateau remnants and benches have high standard deviations.

Lowland areas tend to have a few steep slopes and many gentle ones; their gradients are positively skewed. In mountain areas, the opposite applies as slopes approach gradients limited by slope stability and ROCK MASS STRENGTH. For example, in the Japanese mountains, on igneous and sedimentary rocks, the mode of gradients becomes sharper as altitude increases. The mode is at 33 to 37 degrees in all three ranges of the Japan Alps (central Honshu) above 1,000 m, up to 2,800 m (Figure 67; Katsube and Oguchi 1999). Mean gradients increase with altitude, to maxima of 32 to 35 degrees above 2,000 m. In the high relief of the north-west Himalaya, on crystalline rocks, gradients range from 0 to 60 degrees, with modes of 33–37 and means of 30–34 degrees (Burbank *et al.* 1996) despite varying uplift and denudation rates. These distributions may reflect a dynamic equilibrium with landsliding removing fractured rock and river gradients increasing to transport this. The similarity to Japan may be deceptive in that averaging over several hundred metres reduces the Himalayan measurements.

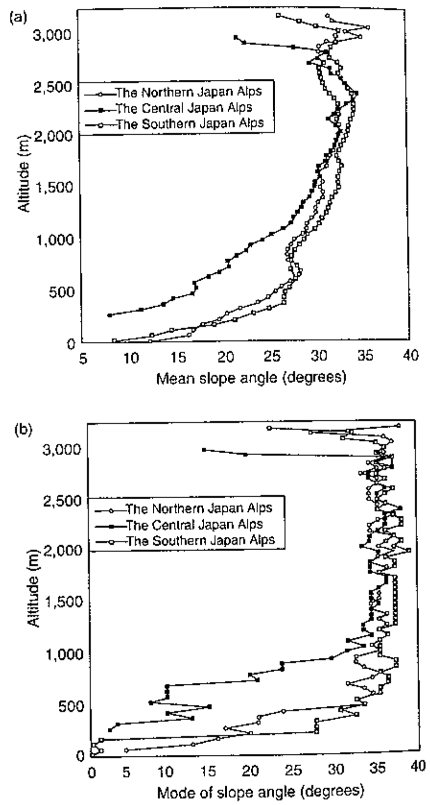


Figure 67 Altitudinal change in (a) mean and (b) modal slope angle (gradient) in three divisions of the Japan Alps

Source: Reproduced from Katsube and Oguchi (1999) with permission from the Association of Japanese Geographers

Gradient is defined as the rate of change of altitude in the direction where that rate is maximized (this is 'true gradient' as opposed to 'apparent gradient' in an arbitrary direction along a profile). This immediately implies a related variable, aspect – the direction or azimuth of this true gradient. Gradient and aspect form a closely related pair, the vector defining surface slope. Aspect, modulated by gradient, has considerable influence on slope climate (mesoclimate), especially solar radiation and exposure to wind. Although slope vectors can be analysed as poles to planes tangential to the surface, that approach ignores the different ways in which gradient and aspect affect surface processes. Aspect is a circular variable (0 ≡ 360 degrees) and it is easy to produce misleading results by applying ordinary linear statistics; it should be summarized by vector, directional or circular statistics, and related to other variables through its sine and cosine, in Fourier Series Analysis.

Rates of change of gradient and aspect in turn define components of curvature, the second derivative of the surface. Evans (1980) defined profile convexity as rate of change of gradient (with negative values representing concavity) and followed earlier geomorphologists such as Young (1972) in expressing this in degrees per 100 m. Plan (contour) convexity is thus the rate of change of aspect. Both variables are encountered in models of surface runoff (see RUNOFF GENERATION). Tangential curvature and other definitions have also been used: mathematically, curvature has three independent components. Of the many ways in which surface curvature can be defined, here we are concerned with those related to the gravity field, which is of central importance in geomorphology. As standard deviation of plan convexity measures the intricacy of contours, it expresses DRAINAGE DENSITY: this relationship requires further investigation.

Surface roughness or ruggedness is a broad concept, covering mean and variability of gradient, and variability of curvature in both profile and plan.

Altitude and its first and second derivatives provide local variables, conceptually related to points although in practice small neighbourhoods are used in their measurement. Context or position on the surface is also important, especially in relation to runoff. Contributing area upslope (per unit width of contour) controls the potential runoff that can be generated, and is used in models and applications (Lane *et al.* 1998; Wilson and Gallant 2000).

Other point aspects of the surface are those which are topologically special, in terms of position: these are summits, saddles and pits, and can be further subdivided in relation to the pattern of higher and lower land in the vicinity. Ridges, valleys and breaks of slope provide linear features at which slope either reverses or changes abruptly. Topological and other linear aspects of the surface are considered under DRAINAGE BASINS. At special points or lines, some derivatives may be indeterminate as gradient passes through zero: notably, aspect and plan convexity/curvature. Plains are areas of zero gradient and again aspect is indeterminate: their extent, however, varies with the vertical resolution of the data (e.g. altitude in metre units, or in tenth-metres, etc.).

General geomorphometry gives an appearance of objectivity, but it involves choice of data source, of horizontal and vertical resolution, and of algorithms for interpolation, smoothing and derivative calculation. Most important of all is definition of areas for which statistical summaries are to be provided. Map sheets or tiles of data are easiest to use, but natural regions may be more appropriate. Islands are the most obvious, but there are two complementary ways in which the land surface may be subdivided into exhaustive, non-overlapping areas. These are drainage basins, and 'mountains' bounded by valleys and low passes.

Spatial series, and complexity

Altitude is a positively autocorrelated variable, that is it defines a generally smooth surface. The rate of decline of autocorrelation with separation is thus an important property, and forms a basis for spectral analysis (Pike and Rozema 1975). This relates to the use of geostatistics and FRACTALS. They provide highly simplified models poorly suited to subaerial topography.

The land surface is complex and its morphometry varies from area to area with rock type and structure, climatic variables and their history, and tectonic history. Attempts to compress its variability into two or three statistical dimensions meet with difficulties. Multivariate studies show that at least nine dimensions (Table 21) are largely independent of each other.

Specific geomorphometry

Taking measurements of landforms requires their precise definition (what is/is not. . .?) and

Table 21 Statistical dimensions of (a) the Wessex land surface, England for 53 areas, and (b) the French land surface, for 72 areas

Property	Statistical descriptor (key variable)	Dimension
	(a) Wessex (Evans)	*(b) France (Depraetere)*
Gradient	Mean gradient	1. Relief
Massiveness	Skewness of altitude	4. Skewness of altitude (and 5.)
Level	Mean altitude	* in 1.
Profile convexity	Skewness of profile convexity	2. Convexity, cols and depressions
Orientation	Weighted vector strength (modulo 180°)	–
Plan convexity	Standard deviation of plan convexity	* in 1.
Altitude-convexity	Correlation of altitude with profile convexity	3. Convexity, crests and slopes
(Profile) variability	Standard deviation of gradient	*in 1.
Directedness	Weighted vector strength (modulo 360°)	–
		5. Skewness of gradient

Notes: All areas 10 × 10 km and analysed from 50 m grids. Numbers in (b) give the rank order of factors
Source: From Evans, in Hergarten and Neugebauer 1999

complete delimitation by a closed outline; here it may be difficult to achieve consistency between researchers. Although specific geomorphometry has been more subjective than general geomorphometry, work has now started on recognition and delimitation of landforms on DEMs by objective criteria. In specific geomorphometry, variables are defined specifically for each landform type. Commonly these include size (length, width, height, area, volume), gradient and shape (often ratios between size variables). The number of possible indices is increased where landforms are subdivided into several parts, e.g. volcano (or impact feature) outer slopes, craters and central peaks. Position (often a surrogate for climate) and geology are sometimes included as potential controlling variables. The more definable landforms include those listed at the end. Each of these has a body of geomorphometric literature. Landform shape and spatial pattern (position relative to others of the same type) were discussed by Jarvis and Clifford (in Richards 1990).

Evans (1987) distinguished eight stages in a specific morphometric study: conceptualization; definition; delimitation; measurement; calculation of indices; analysis of statistical frequency distributions; mapping and spatial analysis; interrelation of attributes; and assessing meaning. Analysis can be in terms of distributions of point variables (altitude, slope and curvature, as discussed above), sets of indices or measurements characterizing each landform (the most common approach), or fitting equations to the whole form or a selected part, outline or profile. In that residuals from such equations usually exhibit spatial pattern, simple equations are rarely good models for landforms.

General concepts in specific geomorphometry include symmetry (radial or axial), scale and the relation of size to shape. The latter can be isometric (shape does not vary with size, expected values of all ratios remain the same) or allometric (see ALLOMETRY; shape changes systematically, often as a power function of size). Scale is fundamental to both general and specific geomorphometry (Dietrich and Montgomery 1998; Wood 1996). Most landforms are defined with specific scales in mind, usually with something like a tenfold range in linear size. Within a particular landform type, different attributes (size, gradient) scale smoothly with each other. Sometimes scale breaks are discovered, and these reveal process thresholds – as for the central features of impact craters (Figure 68; Pike 1980).

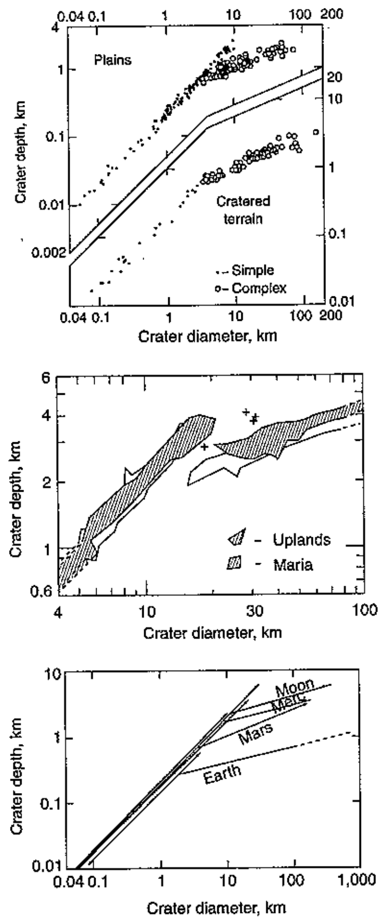


Figure 68 Breaks in the crater depth:diameter scaling relation, illustrating the morphologic transition from simple to complex craters (a) 230 craters on Mars, showing larger simple craters on plains than on 'cratered terrain'; (b) based on 203 mare craters and 136 upland craters on the moon. Simple craters follow a similar relation for maria and for uplands (as for the two divisions of Mars), but complex craters average 12 per cent deeper in uplands; (c) summary of the relationships on three planets and the moon. The transition size increases as gravity decreases

Source: Reproduced from Pike (1980: figures 6, 9 and 2) with permission

References

Burbank, D.W., Leland, J., Fielding, E., Anderson, R.S., Brozovic, N., Reid, M.R. and Duncan, C. (1996) Bedrock incision, rock uplift and threshold hillslopes in the northwestern Himalayas, *Nature* 379, 505–510.

Dietrich, W.E. and Montgomery, D.R. (1998) Hillslopes, channels and landscape scale, in G. Sposito (ed.) *Scale Dependence and Scale Invariance in Hydrology*, 30–60, Cambridge: Cambridge University Press.

Evans, I.S. (1972) General geomorphometry, derivatives of altitude, and descriptive statistics, in R.J. Chorley (ed.) *Spatial Analysis in Geomorphology*, 17–90, London: Methuen.

——(1980) An integrated system of terrain analysis and slope mapping, *Zeitschrift für Geomorphologie N.F. Supplementband* 36, 274–295.

——(1987) The morphometry of specific landforms, in V. Gardiner (ed.) *International Geomorphology 1986* Part II, 105–124, Chichester: Wiley.

Hergarten, S. and Neugebauer, H.J. (eds) (1999) *Process Modelling and Landform Evolution*, Lecture Notes in Earth Sciences, 78, Berlin: Springer.

Katsube, K. and Oguchi, T. (1999) Altitudinal changes in slope angle and profile curvature in the Japan Alps: a hypothesis regarding a characteristic slope angle, *Geographical Review of Japan B* 72, 63–72.

Lane, S.N., Richards, K.S. and Chandler, J.H. (eds) (1998) *Landform Monitoring, Modelling and Analysis*, Chichester: Wiley.

Ohmori, H. (1993) Changes in the hypsometric curve through mountain building and denudation, *Geomorphology* 8, 263–277.

Pike, R.J. (1980) Control of crater morphology by gravity and target type: Mars, Earth, Moon, *Proceedings, Lunar and Planetary Science Conference* 11, 2,159–2,189.

Pike, R.J. (2000) Geomorphometry – diversity in quantitative surface analysis, *Progress in Physical Geography* 24, 1–20.

Pike, R.J. and Dikau, R. (eds) (1995) Geomorphometry, *Zeitschrift für Geomorphologie N.F. Supplementband* 101.

Pike, R.J. and Rozema, W.J. (1975) Spectral analysis of landforms, *Annals of the Association of American Geographers* 64, 499–514.

Richards, K.S. (ed.) (1990) Form, in A. Goudie (ed.) *Geomorphological Techniques*, 31–108, London: Unwin Hyman.

Wilson, J.P. and Gallant, J.C. (eds) (2000) *Terrain Analysis: Principles and Applications*, New York: Wiley.

Wood, J. (1996) Scale-based characterization of digital elevation models, in D. Parker (ed.) *Innovations in GIS* 3, 163–175, London: Taylor and Francis.

Young, A. (1972) *Slopes*, Edinburgh: Oliver and Boyd.

SEE ALSO: hillslope, form; hillslope, process; slope, evolution; and the landforms: alluvial fan; atoll; cave; channel, alluvial (hydraulic geometry); cirque, glacial; crater; doline; drumlin; dune, aeolian; fjord; inselberg; karren; lake; landslide; palsa; pingo; river delta; tafoni; tor; volcano; yardang

IAN S. EVANS

GEOSITE

Geosites (synonyms: geotopes, Earth science sites, geoscience sites) are portions of the geosphere that present a particular importance for the comprehension of Earth history. They are spatially delimited and from a scientific point of view clearly distinguishable from their surroundings. More precisely, geosites are defined as geological or geomorphological objects that have acquired a scientific (e.g. sedimentological stratotype, relict moraine representative of a glacier extension), cultural/historical (e.g. religious or mystical value), aesthetic (e.g. some mountainous or coastal landscapes) and/or social/economic (e.g. aesthetic landscapes as tourist destinations) value due to human perception or exploitation. Various groups of geosites are generally specified in the reference literature: structural, petrological, geochemical, mineralogical, palaeontological, hydrogeological, sedimentological, pedological and geomorphological geosites. In the last case, they are also called geomorphological sites or geomorphosites. Some anthropic objects (e.g. mines) are also considered as geohistorical sites. Geosites can be single objects (e.g. springs, lava streams) and larger systems (e.g. river systems, glacier forefields, coastal landscapes). Active geosites allow the visualization of geo(morpho)logical processes in action (e.g. river systems, active volcanoes), whereas passive geosites testify to past processes; in this case, they have a particular patrimonial value as Earth memory (landscape evolution, life history and climate variations).

Geosites may be modified, damaged, and even destroyed, by natural processes and anthropogenic actions. In order to avoid damage and destruction, geosites need conservation. Conservation strategies are generally based on inventories of geosites requiring the development of assessment methods. Assessment is based on criteria such as integrity (whether the object is complete), exemplarity (to what extent the geosite is representative of the geology or geomorphology of a region or country), rarity (in the space of reference or in scientific terms), legibility (whether it is easily visible scientifically), accessibility (for pedagogic activities), vulnerability, paleogeographical value (its contribution to the history of the Earth), aesthetic value, and cultural/historical value. Geodiversity is a criterion used for assessing groups of geosites: geodiversity is higher where there is a concentration of different objects in a given space facilitating visits and protection. Several quantitative or qualitative procedures for evaluation exist in the literature.

Some countries have adopted specific legislation for geosite conservation (e.g. Great Britain has individuated Regionally Important Geological/Geomorphological Sites – RIGS). Generally, geosite conservation is relatively high in developed countries but low in developing countries.

Further reading

Actes du premier symposium international sur la protection du patrimoine géologique, Digne-les-Bains, 11–16 juin 1991, Mém. Soc. Géol. France, N.S., 165, 1994.

Barettino, D., Vallejo, M. and Gallego, E. (eds) (1999) *Towards the Balanced Management and Conservation of the Geological Heritage in the New Millenium*, III International Symposium ProGEO on the Conservation of the Geological Heritage, Madrid: Ed. Sociedad Geológica de España.

O'Halloran, D., Green, C., Harley, M., Stanley, M. and Knill, J. (eds) (1994) *Geological and Landscape Conservation*, Proceedings of the Malvern International Conference 1993, London: The Geological Society.

Wilson, R.C.L. (ed.) (1994) *Earth Heritage Conservation*, London: The Geological Society and The Open University.

SEE ALSO: geodiversity; landscape sensitivity

EMMANUEL REYNARD

GILGAI

A form of micro-relief consisting of mounds and depressions arranged in random to ordered patterns (Verger 1964). There is a great variety of forms and they occur on a range of swelling clay and texture-contrast soils that have thick subsoil clay horizons. They tend to occur on level or gently sloping plains in areas subject to cycles of intense wetting and drying. Gilgai is an Australian aboriginal word meaning 'small waterhole' (Hubble *et al.* 1983) and some seasonal ponding of water does occur in some of the closed depressions of the larger forms.

The mechanisms of gilgai development involves swelling and shrinking of clay subsoils under a severe seasonal climate. A widely adopted hypothesis for their formation is as follows (Hubble *et al.* 1983: 31):

> when the soil is dry, material from the surface and the sides of the upper part of major cracks falls into or is washed into the deeper cracks, so reducing the volume available for expansion on rewetting of the subsoil. This creates pressures which are revealed by heaving of the soil between the major cracks which, once established, tend to be maintained on subsequent drying. This process is repeated, with the result that the subsoil is progressively displaced, a mound develops between the cracks, and the soil surface adjacent to the cracks is lowered to form depressions.

However, some gilgai are linear forms, known colloquially as 'Adams furrows', 'black-men's furrows', 'stripy country' and 'wavy country' (Hallsworth *et al.* 1955). Beckmann *et al.* (1973: 365) see surface runoff and soil heaving as working together to produce such features, particularly on pediment slopes.

In the Kimberley there are individual linear gilgai up to 2 km long and it is possible that in their case aeolian processes have contributed to their development (Goudie *et al.* 1992).

References

Beckmann, G.G., Thompson, C.H. and Hubble, G.D. (1973) Australian landform example no. 22: linear gilgai, *Australian Geographer* 12, 363–366.

Goudie, A.S., Sands, M.J.S. and Livingstone, I. (1992) Aligned linear gilgai in the west Kimberley District, Western Australia, *Journal of Arid Environments* 23, 157–167.

Hallsworth, E.G., Robertson, G.K. and Gibbons, F.R. (1955) Studies in pedogenesis in New South Wales VIII. The 'Gilgai' soils, *Journal of Soil Science* 6, 1–34.

Hubble, G.D., Isbell, R.F. and Nortcote, K-H. (1983) Features of Australian soils, in Division of Soils, CSIRO, *Soils an Australian Viewpoint*, 17–47, Melbourne: Academic Press.

Verger, F. (1964) Mottureaux et gilgais, *Annales de Géographie* 73, 413–430.

A.S. GOUDIE

GIS

A Geographic Information System (GIS) can be defined as a system of hardware and software used for the capture, storage, management, retrieval, display and analysis of geographic data. GIS's have been around since the 1960s, when, independently, initiatives in Canada (the Canadian GIS (CGIS), developed under the direction of Roger Tomlinson within the Canadian Federal Department of Agriculture) and in the United States (the Laboratory for Computer Graphics at Harvard University, established under the direction of Howard Fisher) resulted in the development of computer-based geographic information systems as we now know them (Foresman 1998).

A GIS uses two fundamental data types: spatial data, consisting of points (e.g. a sample location, a spot height), lines (e.g. the bank of a river, a break in a slope), areas (e.g. a drumlin, a drainage basin) and cells or rasters (e.g. a pixel from a satellite image), and attribute or descriptor data, consisting of characteristics associated with the spatial data (e.g. the elevation of the spot height, the area of the drainage basin).

While initial efforts during the 1970s and early 1980s saw the development of many one-off systems, the emergence in the 1980s of dominant players such as Environmental Systems Research Institute (ESRI) and Intergraph produced a shift from building systems to application development. While geomorphologists were among the first to appreciate the power that computers could bring to scientific analyses (e.g. Chorley 1972), a lack of spatial data and appropriate analytical procedures meant that, initially, such systems were not widely used within the scientific community. As well, the widespread acceptance of GIS required time for scientists to learn, apply and review the new technology in light of contemporary research problems. However, the rise in the computing power of personal computers, coupled with an increased availability of spatial data, meant that by the late 1980s GIS had become a tool used by many geomorphologists.

The earliest links between GIS and geomorphological research can be traced back to applications involving digital elevation models. DIGITAL ELEVATION MODELS (DEMs) are the digital representation of elevation, and while contours have traditionally been used to graphically represent topography on printed maps, in a GIS either a regular tessellation such as gridded cells (i.e. a raster representation, typically referred to as a DEM) or an irregular tessellation such as a Triangulated Irregular Network (TIN) (i.e. a vector representation consisting of elevation points and connecting lines forming triangular planar regions) are the preferred means for representing topography (Weibel and Heller 1991). The main concerns surrounding any digital representation of topography relate to issues around fidelity, accuracy and resolution (Moore *et al.* 1991). While TINs, with their variable sampling structure and their ability to represent important landform features such as peaks, ridges, cliffs and valleys, are capable of more accurately capturing the complexity of topography than are DEMs, DEMs remain the favoured means of representing topography for geomorphologists because of the ease with which analytical procedures can be applied to them (e.g. moving windows of 3 by 3 cells within which attributes such as slope and aspect can be quickly calculated).

The resolution of the data available to the analyst determines the areal extent of the study area that is appropriately analysed, and the type of features that can be identified. Initially, much digital data was of coarse resolution that limited the nature of the analysis (e.g. to regional or macro analyses). As higher resolution data become more commonplace, geomorphologists are better able to study processes at finer spatial and temporal scales, and to examine the role that scale plays in physical processes (Walsh *et al.* 1998).

Extraction of drainage features from any digital representation of topography remains a complex process, however. While with TINs the extraction of the drainage network can be an easy process, determination of the direction and accumulation of surface waters remains a geometrically complex task. For DEMs, the presence of artifactual sinks – either through errors in the elevation values or as a result of the discretization of the elevation values – complicates the process, and much effort has been directed at automatically recognizing and removing such features from DEMs (e.g. Maidment 1993).

Geomorphology is concerned with the form, the materials and the processes from which landforms are created. Some geomorphologists claim that a full understanding of materials and process can be obtained by focusing studies on the form of the landscape (Speight 1974) – a view shared by others in fields such as fractals. GIS are the ideal tools with which to study form, and it is not surprising to find that computer-based morphometric analysis of Digital Elevation Models (DEMs) has long been a very active area of research (e.g. Dikau 1989; Pike 1988; see GEOMORPHOMETRY). Early work focused on the derivation of topographic properties of watersheds, including the derivation of slope, curvature, channel links and drainage areas, all properties that can be mathematically derived from a DEM. This early work focused on examining morphologic patterns, rather than the physical processes that control them. While the link between form and process remains tentative, even at present, we are seeing an increasing sophistication in the application of quantitative approaches to the study of landform, along with the integration of more complex process models within Geographic Information Systems.

The use of GIS in geomorphology can be conceptually classified into four general types of analyses (Vitek *et al.* 1996; see also Walsh *et al.* 1998). These include: (1) landform measurement, (2) landform mapping, (3) process monitoring, and (4) landscape and process modelling. Often an application will involve several levels of GIS analysis. Process monitoring and landscape modelling, for example, will routinely require landform measurement and mapping in order to define initial parameters and establish the spatial distribution of controlling factors of interest. Although the utilization of GIS is not always required for carrying out these types of geomorphological analyses, given suitable digital data and processing capabilities, a GIS can provide a platform for automating these functions. This automation can greatly enhance the spatial/temporal scope and resolution of many geomorphological investigations. A general description and research example for each of the four different types of GIS utilization in geomorphology is provided below.

At the most basic level, a GIS may be used to perform fundamental measurements of landform features. This includes the enumeration of landform features, and making measurements of landform length, area and volume. Often it is some relation between fundamental measures that is of interest. Some common geomorphological examples include counting the number of landslides per region area (event frequency), measuring the length of channels per unit area (drainage density), taking the ratio of horizontal to vertical hillslope or channel lengths (slope), and calculating the areal extent of glaciers in an alpine catchment (glacial coverage). Carrying out these types of landform measurements is usually easily and efficiently accomplished using most standard GIS applications. By automating such measurements in a GIS environment, the spatial scope and resolution of the analysis can be expanded beyond what could be accomplished using traditional manual techniques. For example, Fontana and Marchi (1998) used this type of GIS analysis in order to evaluate

the intensity of localized erosional processes in two alpine drainage basins of the Dolomite Mountains in northeastern Italy. In their work the combination of two landscape measurements of erosion potential was considered – the contributing drainage area (indication of flow concentration occurrence) and local slope (index of flow erosivity). By using the automated measurement capabilities of the GIS, these hydrologic parameters were calculated over entire drainage basin areas of many square kilometres using a high-resolution ten by ten square metre grid base. This type of analysis provided highly localized information on sediment erosion potential within the alpine catchments, information that would have been impossible to obtain using manual methods.

The second way in which GIS is utilized in geomorphology is in the development and analysis of landform maps (see TERRAIN EVALUATION; GEOMORPHOLOGICAL MAPPING). Landform mapping is commonly used in geomorphology in order to characterize landscapes and relate landform distribution to spatial patterns of physical, chemical and biological geomorphic processes. Through the use of basic GIS-based mapping tools, users can rapidly produce and modify landform maps. The typical map manipulation and analysis functions that are included in most GIS applications further enhance landform map production and interpretation. Such functions include map generalization and simplification, map overlay, spatial query and browsing, and various algorithms for the analysis of spatial patterns and relationships. Computer automation is becoming increasingly necessary in landform mapping because of the large amount of digital geographic data that is available and the rapidly increasing rate of digital geographic data acquisition. Bishop *et al.* (1998) used GIS-based mapping for studying large-scale geomorphic processes acting on the Nanga Parbat Himalaya massif of northern Pakistan. A digital elevation model and multispectral remote sensing data were used to study the structural geology and surface geomorphology of this extensive and remote mountain environment. Landform maps, a major component of the investigation, were developed using GIS software and the integration of some massive digital spatial data sets comprised of both surface and subsurface remote sensing data. Assessments of denudation rates and sediment storage were quantified and glacial, fluvial and mass movement processes were reconstructed for the massif by analysing the three-dimensional form and spatial surface characteristics of the mapped landscape in this study. Walsh *et al.* (1998) and Bishop *et al.* (1998) both stress the growing importance of the integration of remote sensing and GIS spatial analysis in order to solve complex, large-scale geomorphic problems.

The analytic capabilities of GIS are well suited to studies of river channel dynamics. At a fundamental level, GIS is an efficient tool for mapping channel features including, for example, sand and gravel bars, vegetated islands, channel banks, woody debris jams and historic (abandoned) channels. Commonly, channel features are digitized from existing maps, orthophotos, aerial photographs or satellite images and coded in the GIS as line or polygon features. Data collected in the field may also be included. If a scale or co-ordinate system has been defined, GIS is an efficient tool for quickly examining spatial relations between, and for measuring the length, width and area of digitized features. If maps or imagery are available for different dates, more advanced spatial and temporal overlay analysis can be performed such as lateral migration, loss or gain of riparian surfaces, aggradation or degradation of sediment, and changes in channel planform. Changes can be examined visually, or summarized and tabulated in a database, while rates of change may be additionally determined if exact dates are known. Similarly, if sufficient historical data are available, GIS may further be used to show trends in morphologic development over time, and even predict landform evolution. The capability of most GIS to collate data sources with different scales and co-ordinate systems is key to these types of analysis. In recent years, GIS has been used increasingly to study the relation between (morphologic) form and (hydraulic) process in river channels using fully distributed topographic models of the channel bed and banks for different dates (cf. Lane *et al.* 1994). Topographic information may be derived from conventional cross-section surveys, tacheometry, photogrammetry or depth soundings. The data are imported to the GIS in order to produce either a TIN or DEM, and then are overlaid in order to produce maps and volumetric summaries of channel scour (erosion) and fill (deposition). The net difference between channel scour and fill may then be used to infer rates of sediment transport within the framework of a sediment budget.

The seamless integration of environmental models with GIS is the ultimate goal of many researchers (Raper and Livingstone 1996). Since landscape and process modelling requires the storing, retrieving and analysing of spatio-temporal data sets, it is not surprising that GIS can play a significant role in this area. GIS enhances the process as it assists in the derivation, manipulation, processing and visualization of such geo-referenced data. Boggs *et al.* (2000) used an integrated approach to look at landform evolution in a catchment that could be subject to impacts from mining activities. Landform evolution models typically require extensive parameterization, often involving both hydrology and sediment transport models, and using an integrated environment allows for the rapid production of modified input scenarios. Therefore, a much wider array of impact scenarios can be made, and the environmental implications of decisions made today can be modelled over the long term, which should lead to better management decisions.

References

Bishop, M.P., Shroder Jr, J.F., Sloan, V.F., Copland, L. and Colby, J.D. (1998) Remote sensing and GIS technology for studying lithospheric processes in a mountain environment, *Geocarto International* 13(4), 75–87.

Boggs, G.S., Evans, K.G., Devonport, C.C., Moliere, D.R. and Saynor, M.J. (2000) Assessing catchment-wide mining-related impacts on sediment movement in the Swift Creek catchment, Northern Territory, Australia, using GIS and landform-evolution modelling techniques, *Journal of Environmental Management* 59, 321–334.

Chorley, R.J. (ed.) (1972) *Spatial Analysis in Geomorphology*, New York: Harper and Row.

Dikau, R. (1989) The application of a digital relief model to landform analysis in geomorphology, in J. Raper (ed.) *Three Dimensional Applications in Geographic Informations Systems*, 51–77, London: Taylor and Francis.

Fontana, G.D. and Marchi, L. (1998) GIS indicators for sediment sources study in alpine basins, in K. Kovar, U. Tappeiner, N. Peters and R. Craig (eds) *Hydrology, Water Resources and Ecology in Headwaters*, 553–560, IAHS Publication no. 248.

Foresman, T.W. (ed.) (1998) *The History of Geographic Information Systems, Perspectives from the Pioneers*, Upper Saddle River, NJ: Prentice Hall.

Lane, S.N., Chandler, J.H. and Richards, K.S. (1994) Developments in monitoring and modelling small-scale river bed topography, *Earth Surface Processes and Landforms* 19, 349–368.

Maidment, D.R. (1993) GIS and hydrological modelling, in M.F. Goodchild, B.O. Parks and L.T. Steyaert (eds) *Environmental Modeling with GIS*, Chapter 14, New York: Oxford University Press.

Moore, I.D. Grayson, R.B. and Ladson, A.R. (1991) Digital terrain modelling: a review of hydrological, geomorphological and biological applications, *Hydrological Processes* 5, 3–30.

Pike, R. (1988) The geometric signature: quantifying landslide-terrain types from digital elevation models, *Mathematical Geology* 20, 491–511.

Raper, J. and Livingstone, D. (1996) High-level coupling of GIS and environmental process modeling, in M. Goodchild, L.T. Steyart, B.O. Parks, C. Johnston, D. Maidment, M. Crane and S. Glendinning (eds) *GIS and Environmental Modeling: Progress and Research Issues*, Fort Collins, CO: GIS World, Inc.

Speight, J.G. (1974) A parametric approach to landform regions, *Institute of British Geography Special Publication*, No. 7, 213–230.

Vitek, J.D., Giardino, J.R. and Fitzgerald, J.W. (1996) Mapping geomorphology: a journey from paper maps, through computer mapping to GIS and virtual reality, *Geomorphology* 16, 233–249.

Walsh, S.J., Butler, D.R. and Malanson, G.P. (1998) An overview of scale, pattern, process relationships in geomorphology: a remote sensing and GIS perspective, *Geomorphology* 21, 183–205.

Weibel, R. and Heller, M. (1991) Digital terrain modelling, in D.J. Maguire, M.F. Goodchild and D.W. Rhind (eds) *Geographic Information Systems: Principles and Applications* Vol. 1, 269–297, Harlow: Longman Scientific.

BRIAN KLINKENBERG, ERIK SCHIEFER AND DARREN HAM

GLACIAEOLIAN (GLACIOAEOLIAN)

The association between glaciation (past and present) and aeolian processes and forms. Glaciation comminutes rock fragments by grinding and so produces material, including silt and sand, that can then be transported by wind to create dunes and LOESS deposits. Such materials are also available to cause wind abrasion in glacial and near-glacial environments, thereby producing wind-moulded pebbles and cobbles (VENTIFACTS) and streamlined ridges (YARDANGS) and grooves. Deflation from glacial deposits can create STONE PAVEMENTS (Derbyshire and Owen 1996).

The fine materials blown across proglacial plains and beyond create dunes and sandsheet deposits (coversands). These are widespread in Canada and Central Europe. The facies progression from coversands through to sandy loess and then loess is well documented from the proglacial forelands around the northern hemisphere. Above all, however, the thick loess deposits of Central Europe, Central Asia, China, New Zealand, the Argentinian Pamas and the mid-USA, may at least in part be indirect products of glacial sediment supply.

Ice sheets and glaciers may themselves affect air pressure conditions and generate high velocity winds that contribute to the power of aeolian processes in their vicinity.

Reference

Derbyshire, E. and Owen, L.A. (1996) Glacioaeolian processes, sediments and landforms, in J. Menzies (ed.) *Past Glacial Environments: Sediments, Forms and techniques, Vol. 2, Glacial Environments*, 213–237, Oxford: Butterworth-Heinemann.

A.S. GOUDIE

GLACIAL DEPOSITION

Glacial deposition occurs when debris is released from glacial transport at the margin or the base of a GLACIER. Narrow definitions include only primary sedimentation directly from the ice into the position of rest, but broader definitions include secondary processes such as deposition through water and resedimentation of glacigenic materials by flowage. In addition, release of material onto the glacier surface is sometimes referred to as supraglacial deposition, but if the glacier is still moving this is only a temporary stage in the sediment transport process.

Glacial deposition produces characteristic sediments and landforms, and landscapes of glacial deposition exist across large areas of the mid-latitudes formerly covered by ice sheets. The term till is generally used to describe sediments deposited by glaciers, and replaces the term boulder clay, which was commonly used in the past. Landforms created by glacial deposition are called MORAINES. Glacial sediments are often unstable in non-glacial environments and are subject to reactivation under PARAGLACIAL conditions. Glacially deposited materials are therefore an important debris source in geomorphic processes in postglacial and proglacial environments.

The characteristics of glacial sediments reflect both the processes of their deposition and also the processes of GLACIAL EROSION and entrainment by which the material was originally produced. A substantial literature exists on the classification of glacial sediments and the processes by which they form (e.g. Schlüchter 1979; van der Meer 1987; Goldthwait and Matsch 1988) and several convenient summaries of depositional processes have been produced (e.g. Whiteman 1995). Following a broad definition, the main mechanisms of deposition include:

1 release of debris by melting or sublimation of the surrounding ice
2 lodgement of debris by friction against a substrate
3 deposition of material from meltwater (glacifluvial deposition)
4 chemical precipitation
5 flow and resedimentation of deposited material
6 glacitectonic processes (see GLACITECTONICS).

Different processes of sedimentation are dominant in different parts of a glacier. In supraglacial locations, material can be released by ablation of the ice surface. This occurs most commonly by melting, and is referred to as melt-out, but sublimation can make a significant contribution in cold arid environments (Shaw 1988). Englacial debris, and debris in basal ice penetrating to the surface, is then exposed on the ice surface and can contribute to a supraglacial sediment layer. This supraglacial sediment is liable to redistribution by flow, wash and mass movements on the ice surface as ablation continues. Resedimented material derived from flow of supraglacial debris, sometimes referred to as 'flow-till', can make up a substantial proportion of glacial deposits in environments where supraglacial sedimentation occurs (e.g. Boulton 1968). The extent of supraglacial sedimentation depends on the debris content of the ice, the ablation rate and the rate of removal of sediment from the surface by processes such as wash, deflation and mass movement. Supraglacial sediment can be deposited at the ground surface when the glacier retreats or disintegrates.

At glacier margins material can be released by ablation and dumped directly into the proglacial environment. Sediment can be released directly from within the ice, carried to the margin supraglacially and dropped over the margin as if over the end of a conveyor belt, or brought to the margin by water flowing from the interior or surface of the glacier. Sediment can also be transferred to the margin by deformation of subglacial material (e.g. Boulton *et al.* 1995). The amount of sediment that is transported to the margin is primarily a function of the size and speed of the glacier, the glacier's erosive capability, the erodibility of the substrate and the input of sediment from extra-glacial sources such as rockfalls or tephra. The characteristics of the sediments and landforms produced by deposition at the margin depend on whether the margin occurs on land or in water, and the processes and environments of their formation are reflected in their morphology and sedimentological structure. Reactivation and resedimentation of material at the margin by movement of the ice or by fluvial and mass movement processes is common.

In subglacial environments, material can be released by ablation either in cavities or in contact with the bed, or can be lodged against the bed by moving ice. Release of basal sediment by melt-out or sublimation beneath moving ice can produce conditions conducive to lodgement, to the development of thick basal till, and to subglacial deformation of the released material. Lodgement of sediment against a rigid bed occurs when the friction of the clast against the bed outweighs the tractive power of the ice and material is released from the ice by either pressure melting or plastic deformation of ice around clasts. Lodgement can also occur against the upper surface of a deforming bed if the overlying ice is moving faster than the deforming layer. Deposition within a deforming subglacial layer occurs when deformation cannot remove all of the material that is being supplied to the layer and deformation ceases either throughout the layer or for a certain thickness at its lower margin.

Subglacial deposits can also include chemical precipitates, although these are not always considered in the context of glacial sediments. Chemical precipitates such as calcite can be deposited when solute-rich waters freeze. In carbonate environments, such as where limestone bedrock is present, comminuted carbonate rocks contribute to a highly reactive rock flour. Meltwater produced at the bed can take carbonate into solution, and if the water subsequently refreezes the carbonate can be released as a precipitate onto bedrock or basal debris (e.g. Souchez and Lemmens 1985).

Many glaciers release sediment into water, and GLACIMARINE and GLACILACUSTRINE sediments form an important part of the glacial sediment record. Release of sediment into water can produce different effects from terrestrial sedimentation. Sediment can be released directly from the ice, by discharge of sediment-rich meltwater, and by the melting or breakup of icebergs. The characteristics of subaqueous sediments and landforms reflect aqueous as well as glacial processes and conditions. Where a glacier is grounded on its bed beneath the water, glacial deposition can occur beneath the margin by lodgement and melt-out. However, material that is dumped from the front of the glacier into water, or from the base of the glacier where the glacier is floating, forms a deposit that is not strictly a glacial deposit, as its character upon settling will be controlled largely by aqueous sedimentation processes. The principal features of ice margins in water include subaqueous moraines caused both by pushing of proglacial sediments and release of sediment from the glacier; subaqueous grounding line fans formed from material emerging from beneath the glacier into the water at the grounding line; ice contact fan deltas that form when grounding line fans grow and emerge at the water surface; and a distal proglacial zone in which sediment settles out from suspension in the water and rains out from icebergs drifting away from the ice margin. Useful reviews of sedimentation at marine and lacustrine glacier margins include those provided by Dowdeswell and Scourse (1990) and Powell and Molnia (1989).

The mechanisms of glacial deposition impart specific characteristics to the deposited material. Fabric, particle size and shape characteristics, and consolidation have all been used to infer depositional processes and glacier characteristics from glacial sediments. The distribution of structures in deformed sediments can be used to reconstruct former ice sheets, and Boulton and Dobbie (1993) suggested that consolidation characteristics of formerly subglacial sediments can be used to infer basal melting rates, subglacial groundwater flow patterns, ice overburden, basal shear stress, ice-surface profiles and the amount of sediment removed by erosion. Glacial deposits can thus provide valuable information about glacial and climatic history.

References

Boulton, G.S. (1968) Flowtills and related deposits on some Vestspitsbergen Glaciers, *Journal of Glaciology* 7, 391–412.

Boulton, G.S., and Dobbie, K.E. (1993) Consolidation of sediments by glaciers: relations between sediment geotechnics, soft-bed glacier dynamics and subglacial ground-water flow, *Journal of Glaciology* 39, 26–44.

Boulton, G.S., Caban, P.E. and van Gijssel, K. (1995) Groundwater flow beneath ice sheets: part I – large scale patterns, *Quaternary Science Reviews* 14, 545–562.

Dowdeswell, J.A. and Scourse, J.D. (1990) *Glacimarine Environments: Processes and Sediments*, Geological Society Special Publication 53, Bath: Geological Society.

Goldthwait, R.P. and Matsch, C. (1988) *Genetic Classification of Glaciogenic Deposits*, Rotterdam: Balkema.

Powell, R.D. and Molnia, B.F. (1989) Glacimarine sedimentary processes, facies and morphology of the south-southeast Alaska shelf and fjords, *Marine Geology* 85, 359–390.

Schlüchter, C. (ed.) (1979) *Moraines and Varves*, Rotterdam: A.A. Balkema.

Shaw, J. (1988) Sublimation till, in R.P. Goldthwait and C. Matsch (eds) *Genetic Classification of Glaciogenic Deposits*, 141–142, Rotterdam: A.A. Balkema.

Souchez, R.A. and Lemmens, M. (1985) Subglacial carbonate deposition: an isotopic study of a present-day case, *Palaeogeography, Palaeoclimatology, Palaeoecology* 51, 357–364.

van der Meer, J.J.M. (ed.) (1987) *Tills and Glaciotectonics*, Rotterdam: A.A. Balkema.

Whiteman, C.A. (1995) Processes of terrestrial deposition, in J. Menzies (ed) *Modern Glacial Environments*, 293–308, Oxford: Butterworth-Heinemann.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.

Bennett, M.R. and Glasser, N.F. (1996) *Glacial Geology*, London: Wiley.

Hambrey, M.J. (1994) *Glacial Environments*, London: UCL Press.

Knight, P.G. (1999) *Glaciers*, Cheltenham: Nelson Thornes.

SEE ALSO: glacier; moraine

PETER G. KNIGHT

GLACIAL EROSION

Glaciers cause erosion in a variety of ways (Bennett and Glasser 1996). First of all, glaciers can be likened to conveyor belts. If a rockfall puts coarse debris on to a glacier surface, for example, or if frost-shattering sends down a mass of angular rock fragments on to the glacier surface, it can then be transported, almost whatever its size, down valley. Second, beneath glaciers there is often a very considerable flow of meltwater. This may flow under pressure through tunnels in the ice at great speed, and may be charged with coarse debris. Such subglacial streams are highly effective at eroding the bedrock beneath a glacier. This can contribute to the excavation of TUNNEL VALLEYs. Meltwater may also cause chemical erosion. Third, although glacier ice itself might not cause marked erosion of a rock surface by *abrasion*, when it carries coarse debris at its base some abrasion can occur. This grinding process has been observed directly by digging tunnels into glaciers, but there is other evidence for it: rock beneath glaciers may be *striated* or scratched, and much of the debris in glaciers is ground down to a fine mixture of silt and clay called *rock flour*.

Glaciers also cause erosion by means of *plucking*. If the bedrock beneath the glacier has been weathered in preglacial times, or if the rock is full of joints, the glacier can detach large particles of rock. As this process goes on, moreover, some of the underlying joints in the rock may open up still more as the overburden of dense rock above them is removed by the glacier. This is a process called *pressure release*.

As debris-laden ice grinds and plucks away the surface over which it moves, characteristic landforms are produced which give a distinctive character to glacial landscapes. Of the features resulting from glacial quarrying, one of the most impressive is the CIRQUE. This is a horseshoe-shaped, steep-walled, glaciated valley head. As cirques evolve they eat back in the hill mass in which they have developed. When several cirques lie close to one another, the divide separating them may become progressively narrowed until it is reduced to a thin, precipitous ridge called an *arête*. Should the glaciers continue to whittle away at the mountain from all sides, the result is the formation of a pyramidal *horn*.

With valley glaciation the lower ends of spurs and ridges are blunted or truncated; the valleys assume a U-shaped configuration; they become more linear; and hollows or troughs are excavated in their floors. Many high-latitude coasts, such as those of Norway, New Zealand and western Scotland, are flanked by narrow troughs, called FJORDS, which differ from land-based glacial valleys in that they are submerged by the sea.

Fjards are related to fjords. They are coastal inlets associated with the glaciation of a lowland coast, and therefore lacking the steep walls characteristic of glacial troughs. A good example of a fjard coast is that of Maine, USA.

A further erosional effect of valley glacier is the breaching of watersheds, for when ice cannot get away down a valley fast enough – perhaps because its valley is blocked lower down by other ice or because there is a constriction – it will overflow at the lowest available point, a process known as *glacial diffluence*. The result of this erosion is the creation of a *col*, or a gap in the watershed.

Tributary valleys to a main glacial trough have their lower ends cut clean away as the spurs between them are ground back and truncated. Furthermore, the floor of a trunk glacier is deepened more effectively than those of feeders from the side or at the head, so that after a period of prolonged glaciation such valleys are left hanging above the main trough. Such *hanging valleys* have often become the sites of waterfalls.

The development of an ice sheet tends to scour the landscape. In Canada there are vast expanses of territory where the Pleistocene glaciers scoured and cleaned the land surface, removing almost all the soil and superficial deposits and exposing the joint and fracture patterns of the ancient crystalline rocks beneath. Streamlined and moulded rock ridges develop, including *roches moutonnées*. In parts of Scandinavia and New England these may be several kilometres long and have steepened faces of more than 100 m in height. They are interspersed with scoured hollows which may be occupied by small lakes when the ice sheet retreats. In western Scotland relief that is dominated by this mixture of rock ridges and small basins is called *knock and lochan* topography. Areas of calcareous rocks that have been scoured by ice often display LIMESTONE PAVEMENTS.

Considerable debate has been attached to the question of rates of glacial erosion (Summerfield and Kirkbride 1992) and some have argued (see GLACIAL PROTECTIONISM) that glaciers can protect the underlying surface from erosion. Much depends on the context in which a glacier or ice sheet occurs, but even for an area like the Laurentide Shield there is controversy (see Braun 1989).

A range of methods has been used to measure amounts of glacial erosion. These have included:

1. The use of artificial marks on rock surfaces later scraped by advancing ice.
2. The installation of platens to measure abrasional loss.
3. Measurements of the suspended, solutional and bedload content of glacial meltwater streams and of the area of the respective glacial basins.
4. The use of sediment cores from lake basins of known age which are fed by glacial meltwater.
5. Reconstructions of preglacial or interglacial land surfaces.
6. Estimates of the volume of glacial drift in a given region and its comparison with the area of the source region of that drift.
7. Cosmogenic nuclides (Colgan *et al.* 2002).

Many published rates of glacial erosion are high, typically 1,000–5,000 $m^3 km^{-2} a^{-1}$, but they vary greatly according to the measuring techniques that are used (Warburton and Beecroft 1993) and to the nature of the environment (Embleton and King 1968). Clearly, geographical location is an important control of the rate of glacial erosion. Some areas have characteristics that limit the power of glacial erosion (e.g. resistant lithologies, low relief, frozen beds), but other areas suffer severe erosion (e.g. non-resistant lithologies, proximity to fast ice streams, thawed beds, etc.).

References

Bennett, M.R. and Glasser, N.F. (1996) *Glacial Geology: Ice Sheets and Landforms*, Chichester: Wiley.

Braun, D.D. (1989) Glacial and periglacial erosion of the Appalachians, *Geomorphology* 2, 233–256.

Colgan, P.M., Bierman, P.R., Mickelson, D.M. and Caffree, M. (2002) Variation in glacial erosion near the southern margin of the Laurentide ice sheet, south-central Wisconsin, USA; implications for cosmogenic dating of glacial terrains, *Geological Society of America Bulletin* 1,114, 1,581–1,591.

Embleton, C. and King, C.A.M. (1968) *Glacial and Periglacial Geomorphology*, London: Arnold.

Summerfield, M.A. and Kirkbride, M.P. (1992) Climate and landscape response, *Nature* 355, 306.

Warburton, J. and Beecroft, I. (1993) Use of meltwater stream material loads in the estimation of glacial erosion rates, *Zeitschrift für Geomorphologie*, 37, 19–28.

A.S. GOUDIE

GLACIAL ISOSTASY

'ISOSTASY' (= equal standing) refers to an equal distribution of pressure or mass within a column of rock that extends from the surface of the Earth to its interior. This concept of equilibrium is, however, perturbed by forces at the surface and within the Earth which create unequal mass distributions. With time, such inequalities are compensated for by adjustments within the lithosphere and aesthenosphere, and the rate and relaxation time for this recovery to equilibrium is a function of the Earth's rheology. The term 'glacial isostasy' is thus used to define the adjustments of the Earth's surface to the growth and disappearance of ice sheets (Andrews 1974). For example, 22,000 years ago the surface of the Earth differed fundamentally from today's geography, as large ice sheets, several kilometres thick, covered much of Canada and most of the area centred on the Baltic (the Fennoscandian ice sheet), large areas of Great Britain, and many

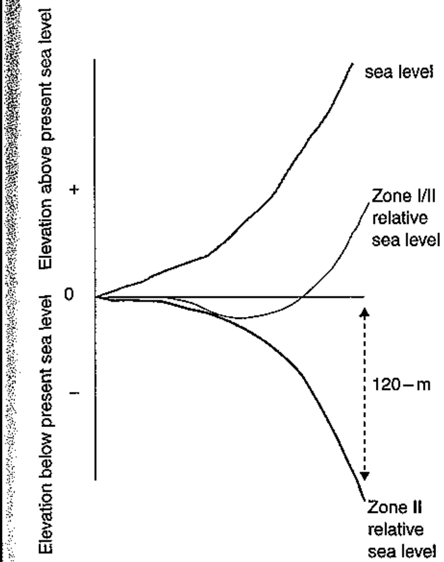


Figure 69 Schematic depiction of changes of relative sea level at sites within the borders of former ice sheets (Zone I), at sites distant (>1,000 km) from the maximum ice extent (Zone II), and sites in the transition between Zones I and II

other parts of the world (Denton and Hughes 1981). Together these ice sheets and glaciers extracted about 120 m of water from the global ocean, thus reducing the water load on the seafloor. During the last 15,000 years or so the ice sheets melted, resulting in a reduction in the load of ice sheets on continents, but the resulting meltwater has added to the load over ocean basins (Peltier 1980). Thus 'glacial isostasy' in its most complete definition includes the global-wide changes in the differential elevation of land due to ice sheet removal, and changes in relative sea level around all the world's coastlines caused by the unique combination of ice and water loading and/or unloading at each location (Figure 69, Zones I, I/II, and II) (Clark *et al.* 1978).

Observations on changes in sea level within formerly glaciated areas have been made for over 150 years, but the link between the removal of the ice load, crustal recovery and isostasy were made in north-west Europe and North America by the mid to late nineteenth century (Andrews 1974). Although significant research was carried out in the early part of the twentieth century studies of glacial isostasy bloomed in the period after ~1960 due in no small part to the development of radiocarbon dating, and to increased levels of field research in Arctic Canada, Greenland and Svalbard. In these areas materials to date the changes of sea level through time (molluscs, whalebone and driftwood) were abundant. The combination of a technique (^{14}C dating) plus exploration of vast tracts of formerly glaciated areas resulted in an explosion of data on changes in sea level and on the delimitation of former ice sheet margins through time (Andrews 1970; Blake 1975; Dyke 1998; Dyke and Peltier 2000; Forman *et al.* 1995). In the mid- to late 1970s these data-rich field observations attracted the interests of the geophysicists who now had both the mathematical tools and the computer power to tackle what is a global Earth-science problem with many ramifications (Peltier and Andrews 1976).

Studies of 'glacial isostasy' represent a significant interplay between workers in several different fields (Figure 70), including (1) the glacial geologist who maps the time-dependant changes in ice sheet extent (and more problematically, thickness); (2) the Quaternary scientist working on changes in sea level from sites within the margins of present ice sheets (near-field sites), to those at and just beyond the former ice sheet limits, and finally to workers reconstructing changes in sea level at sites far-distant (the far-field) from ice margins (say the central Pacific Islands); (3) glaciologists who combine data from (1) above with knowledge of the physics of ice to model changes in ice sheet extent and volume; (4) the geophysicists who 'tune' the behaviour of the Earth's rheology to match the data from inputs (1) and (2). The schematic interplay between these disciplines (Figure 70) has resulted in a series of successive approximations of each of the key components. This process started in 1976 (Peltier and Andrews 1976) and is still continuing.

The rheology of the Earth is most frequently modelled as a self-gravitating viscoelastic (Maxwell) solid (Cathles 1975), although a case can be made for a more complex rheology where the response is a nonlinear (power > 1) function, not unlike the behaviour of glacial ice. A simple 2-D model consists of a lithosphere of some thickness which overlies a fluid aesthenosphere. The lithosphere is rigid with the application of a small

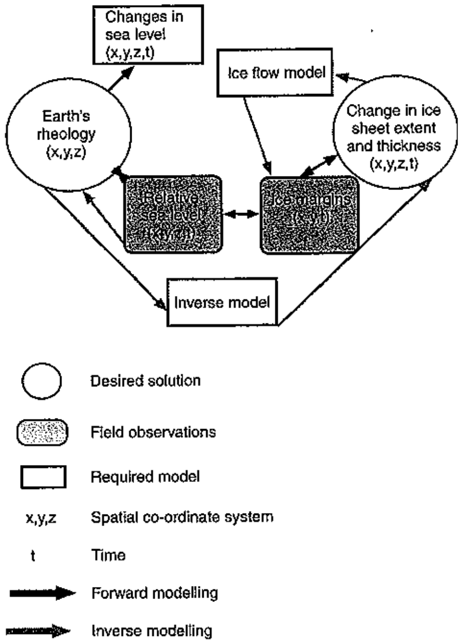


Figure 70 Schematic diagram on interactions between data and models

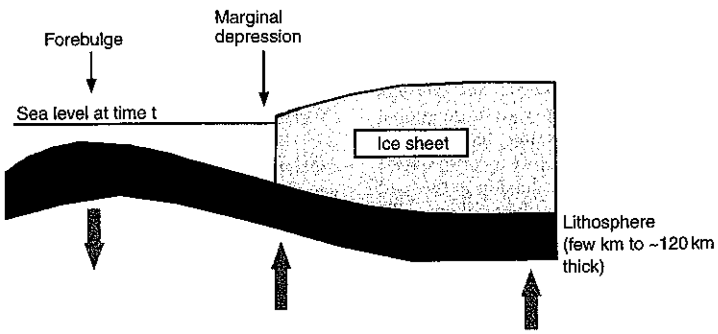


Figure 71 Two-dimensional model of an Earth with a lithosphere and aesthenosphere. The distance between the ice margin and the forebulge is a function of the thickness and elastic properties of the lithosphere. The senses of motion are those during the retreat phase of an ice sheet

load ('small' being a function of the lithosphere thickness, but on old continental shields the load may have to exceed a diameter of ~300 km, whereas on Iceland, with the mantle virtually at the surface, then the load diameter is probably a few kilometres to a few tens of kilometres at most for isostatic compensation by flow to be induced. As an ice sheet grows on land at some point

the lithosphere will bend and material will be displaced by flow. It is generally assumed that the flow can be approximated by a layered Newtonian viscous fluid with a viscosity in the range of 10^{22} poises. The 2-D model of the Earth's response to a glacial load (Figure 71) shows that at the margins of the ice sheet the load is partly supported by the lithosphere so that there is a depression at the ice margin but at some distance from the ice sheet there is a zone of uplift in the forebulge. Upon retreat of the ice sheet the forebulge will collapse, hence there will be a rise in relative sea level.

Figure 70 shows the elements of the problem. Glacial isostasy has been examined in terms of 'forward models' and 'inverse models' in a full 3-D global model. In reality there is an ongoing iteration between the field scientists and the modellers so that both approaches are required (Lambeck 1995; Lambeck *et al.* 1998; Peltier 1994; Peltier 1996). In the 'forward' case, the explicit data are the positions of the ice sheet margins through time (x, y, t), and changes in relative sea level at a suite of sites from Zones I, I/II, and II. What then has to be approximated is the changes in thickness within an ice sheet (x, y, z, t). The application of this time-dependent load to a model of the Earth's rheology will result in changes in sea level (required model) where the predicted changes in sea level include not only the obvious fall of sea level in Zone I, but also account for the transfer of mass (meltwater) from the melting ice sheets to the oceans. Disagreements between the observed relative sea levels (Figure 70) and the predicted sea levels could be related to either an incorrect Earth rheology or an incorrect estimate of changes in the ice sheets in all four dimensions. In contrast, inverse modelling is an attempt to develop a model of the global ice sheet changes by taking the observed relative sea-level data, assuming a rheology, and then using these data to reconstruct the changes in the ice sheets. Appropriate ice flow models can then be applied to see if the reconstructions are glaciologically feasible and whether the reconstructed ice sheets match the data, in this case the mapped and dated ice sheet margins.

References

Andrews, J.T. (1970) A geomorphological study of post-glacial uplift with particular reference to Arctic Canada, Institute of British Geographers Special Publication No. 1.

Andrews, J.T. (1974) *Glacial Isostasy*, Stroudburg, PA: Dowden, Hutchinson and Ross.

Blake, J.W. (1975) Radiocarbon age determination and postglacial emergence at Cape Storm, Southern Ellesmere Island, Arctic Canada, *Geografiska Annaler* 57, 1–71.

Cathles, L.M. III (1975) *The Viscosity of the Earth's Mantle*, Princeton: Princeton University Press.

Clark, J.A., Farrell, W.E. and Peltier, W.R. (1978) Global changes in postglacial sea level: a numerical calculation, *Quaternary Research* 9, 265–287.

Denton, G.H. and Hughes, T.J. (1981) *The Last Great Ice Sheets*, New York: Wiley.

Dyke, A.S. (1998) Holocene delevelling of Devon Island, Arctic Canada: implications for ice sheet geometry and crustal response, *Canadian Journal of Earth Science* 35, 885–904.

Dyke, A.S. and Peltier, W.R. (2000) Forms, response times and variability of relative sea-level curves, glaciated North America, *Geomorphology* 32, 315–333.

Forman, S.L., Lubinski, D., Miller, G.H., Snyder, J., Matishov, G., Korsun, S. and Myslivets, V. (1995) Postglacial emergence and distribution of late Weichselian ice-sheet loads in the northern Barents and Kara seas, Russia, *Geology* 23, 113–116.

Lambeck, K. (1995) Constraints on the Late Weichselian Ice Sheet over the Barents Sea from observation of raised shorelines, *Quaternary Science Reviews* 14, 1–16.

Lambeck, K., Smither, C. and Johnston, P. (1998) Sea-level change, glacial rebound and mantle viscosity for northern Europe, *Geophysical Journal International* 143, 102–144.

Peltier, W.R. (1980) Models of glacial isostasy and relative sea level, *Dynamics of Plate Interiors* 1, 111–128.

Peltier, W.R. (1994) Ice age paleotopography, *Science* 265, 195–201.

——(1996) Mantle viscosity and ice-age ice sheet topography, *Science* 273, 1,359–1,364.

Peltier, W.R. and Andrews, J.T. (1976) Glacial-isostatic adjustment: I The forward problem, *Geophysical Journal Royal Astronomical Society* 46, 605–646.

JOHN T. ANDREWS

GLACIAL PROTECTIONISM

The belief that the erosive power of rain and rivers far exceeds that of glacier ice, and that the presence of glaciers in a region protects the landscape from much more effective fluvial attack (Davies 1969). Glaciers were thought, following Ruskin, to rest in depressions like custard in a pie dish, rather than to erode the basins. Proponents of this theory included the British geologists J.W. Judd, T.G. Bonney, E.J. Garwood and S.W. Wooldridge.

In some areas, where ice stream velocities are low and relief is limited, the erosive role of glaciers may well be passive rather than active. Some degree of protection may be afforded.

Reference

Davies, G.L. (1969) *The Earth in Decay*, London: Macdonald.

A.S. GOUDIE

GLACIAL THEORY

The belief that in former times glaciers had been more extensive than they are today. Some suggestions as to this had originally been made at the end of the eighteenth century. In 1787 de Saussure recognized erratic boulders of palpably Alpine rocks on the slopes of the Jura Ranges, and Hutton reasoned that such far-travelled boulders must have been glacier-borne to their anomalous positions. Playfair extended these ideas in 1802, but it was in the 1820s that the *Glacial Theory*, as it came to be known, really became widely postulated. Venetz, a Swiss engineer, proposed the former expansion of the Swiss glaciers in 1821, and his ideas were supported and strengthened by Charpentier in 1834. The poet Goethe expressed the idea of 'an epoch of great cold' in 1830. However, the ideas of both Venetz and Charpentier were extended and widely publicized by their fellow countryman, Louis Agassiz, who was one of the originators of the term *Eiszeit* or *Ice Age*. In Norway Esmark put forward similar ideas in 1824, and in 1832 Bernhardi went so far as to suggest that the great German Plain had once been affected by glacier ice advancing from the North Polar region.

In spite of this convergence of opinion from numerous sources, these ideas were not easily accepted or assimilated into prevailing dogma, and for many years it was still believed that glacial till, called drift, and isolated boulders, called erratics, were the result of marine submergence, much of the debris, it was thought, having been carried on floating icebergs. Sir Charles Lyell noted debris-laden icebergs on a sea-crossing to America, and found that such a source of the drift was more in line with his belief in the power of current processes – UNIFORMITARIANISM – than a direct glacial origin.

Even towards the end of the nineteenth century some opposition still remained. In 1892, for instance, H.H. Howorth produced his massive neocatastrophist *The Glacial Nightmare and the Flood – a second appeal to common sense from the extravagance of some recent geology*, and tried to return to a fundamentalist-catastrophic interpretation of the evidence.

Conversely, others were overenthusiastic about the glacial theory and Agassiz himself postulated that glaciers reached the humid tropics in South America.

Further reading

Chorley, R.J., Dunn, A.J. and Beckinsale, R.P. (1964) *History of the Study of Landforms*, Vol. 1, London: Methuen.

Davies, G.L. (1969) *The Earth in Decay*, London: Macdonald.

A.S. GOUDIE

GLACIDELTAIC (GLACIODELTAIC)

The discharge of sediment-laden streams from melting glaciers into lakes or fjords usually results in accumulation of a delta. The morphology and structure of the delta depends on several factors. These include nature of the ice margin; a vertical glacier front calving into a fjord contrasts strongly with a gently sloping ice surface descending into a shallow lake overlying dead ice. The sources of meltwater, flowing directly from the ice or entering the water body at the surface, deeper within or at the base of the ice as overflows, interflows or underflows, are also important. The quantity and particle size of sediment, the channel gradient, and the location of deposition close to or distant from the ice margin all influence the form, structure and sedimentary characteristics of the delta.

Deltas formed by discharge of glacial meltwater are located either in the proglacial environment or in ice-contact situations. In the proglacial environment they are deposited in lakes and fjords, and distinct types referred to as Hjulstrom, Gilbert and Salisbury deltas have been identified. In ice-contact situations deltas may form in supraglacial lakes, subglacial water bodies and lakes of the terminoglacial environment.

Hjulstrom deltas occur either in lakes or fjords where outwash sheets (the Icelandic sandur) of gravel and sand enter shallow water with gently

sloping fronts. They may also form fans or small deltas in lakes on the ice surface. The delta gravels and sands are the flood traction loads of braided sandur channels or of fan channels that on deposition in standing water become smaller in grain size away from the point of discharge. Beyond the delta front, which advances into the water as a series of lobes due to changes in channel discharge points, the suspended silts and clays are deposited as prodelta mud.

The Salisbury type delta is intermediate in type and size between the Hjulstrom and Gilbert-type deltas. It forms where high-energy flow from a subglacial tunnel mouth supplies material rapidly via sheet- and streamfloods, and topset beds accrete very rapidly.

The classic glacigenic delta is the Gilbert type that may be formed on ice, in proglacial lakes and in fjords where the water is deep enough to allow development of a distinct structure that includes bottomset, foreset and topset beds. The bottomset beds occur beneath and beyond the foreset beds and extend as prodelta clays to merge with the lake – or fjord floor clays. The bottomset beds result when the suspended sediment carried in turbid flows beyond the delta front settles. The beds are generally laminated. The laminae reflect grain-size variations between fine sand, silt and clay layers due to deposition that may be controlled by seasons, short periods or single events. The sediments become finer away from the point of discharge. They may contain occasional dropstones derived from floating ice, and convoluted laminations due to disturbance of the delta face by slumping, sediment flowage and turbidity currents. Sediment loading and dewatering will also produce convoluted laminations.

The foreset beds dip steeply (*c*.25–30°) away from the source of the glacial sands and gravels and prograde into the lake by the intermittent avalanching down the delta front of cohesionless debris flows. The foresets decrease slightly in slope and sediment size towards the delta face, and individual foresets tend to decrease in grain size upwards. Channels that result from erosion of the delta face by turbidity flows may be cut into the foreset beds. As the delta builds up to water level the glacifluvial sandur deposits extend onto the surface as topset beds. The topset gravels and sands are coarser than those of the foresets, are gently inclined (2–5°), relatively thin (*c*.1–2 m) and exhibit cut-and-fill structures related to the shifting courses of the braided channels of the sandur.

Small ponds and lakes develop on the decaying marginal and terminal parts of glaciers. Meltwater from surface streams or shallow depth within the ice may form small deltas in the shallow lakes. The delta sediments consist of inclined beds of gravel and sand interbedded with unsorted debris flow sediments from the ice surface. Subsequent melting of the underlying and laterally supporting ice causes slumping and flowage of the sediments with the development of fault and fold structures. Where preserved, the sediments form delta-kames.

Subglacial and englacial streams may enter bodies of standing water at or near the base of the ice. Reduction in stream velocity will cause sedimentation of the sands and gravels to form a delta. Decay of the ice may result in formation of a delta-kame.

More extensive lakes are frequently formed in the ice contact terminoglacial environment of glaciers and ice sheets. Where large meltwater streams enter lakes or fjords, deltas of the Gilbert type are formed. Bottomset beds may be formed where the water body is relatively large but most of the delta sediments consist of foreset beds of sand and gravel. Topset beds only develop where a sandur forms between the meltwater outlet and water body. Interstratification of debris flow deposits and foreset sands and gravels is common. Removal of ice support during decay causes collapse and pitting of the ice proximal delta margin. Such deltas have been described as kamiform and where subglacial stream outlets are numerous along an ice edge many may be developed and may coalesce laterally as a kame moraine. The Salpausselka Moraines formed in southern Finland during the final readvance stage, the Younger Dryas, of the last glaciation are over 600 km in length and largely form delta moraines.

In southern Finland the three major Salpausselka delta moraines – formed in the Baltic Ice Lake before the postglacial rise in sea level drowned the Baltic Sea – record retreat stages of the ice sheet margin. Where deltas are formed sequentially inland along fjord margins the stages of glacier retreat in the valley and relative sea-level rise in the fjord can be detected. When ice barriers impounded glacial lakes in upland valleys and formed deltas and shorelines at a number of water levels, as in Glenroy, Scotland, the delta surfaces and shorelines record the stages in draining of the glacial lake.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.
Bennett, M.R. and Glasser, N.F. (1996) *Glacial Geology: Ice Sheets and Landforms*, Chichester: Wiley.
Brodzikowski, K. and van Loon, A.J. (1991) *Glacigenic Sediments*, Amsterdam: Elsevier.
Drewry, D. (1986) *Glacial Geologic Processes*, London: Arnold.

ERIC A. COLHOUN

GLACIER

Glaciers are accumulations of snow and ice on the Earth's surface. They form predominantly at high latitude (polar regions) and at high elevation (on mountains). Here, two meteorological variables, temperature and precipitation, combine to yield conditions where the annual amount of snowfall (predominantly during the cold or wet season) outweighs the annual amount of snow melt (predominantly during the warm or dry season). Under these conditions, consecutive annual snow layers develop, one on top of the other, the pressures of which force snow at depth to change structure and density (recrystallization). These conditions first produce firn/névé (density of 0.400–0.830 $kg\,m^{-3}$) and, when pores are sealed off to create air bubbles, ice (density of 0.830–0.917 $kg\,m^{-3}$). These changes can occur within one year and at shallow depths in maritime regions or take hundreds to thousands of years and considerable depths in continental locations. Glaciers are said to have formed when glacier flow occurs, that is when ice is thick enough to deform plastically under its own weight (Paterson 1994).

Glacier systematics

Glaciers are commonly ordered using their geomorphological or thermal characteristics, both of which yield a tripartite system. The common characteristic of ICE SHEETS and ice caps (smaller versions of the former, i.e. $<50{,}000\,km^2$) is that they are so thick relative to the landscape relief on top of which they rest, that their surface topography and flow direction are unconstrained by the underlying topography (except near the ice margin). Their surface morphology dominantly shows the presence of ice domes, the dome-shaped central regions where ice forms and flows outward towards its perimeter, and outlet glaciers and ice streams, the narrow flow-parallel bands of faster flowing ice that are the primary routes by which the ice is evacuated in marginal areas (Plate 51A). Two ice sheets and numerous ice caps occur in the polar regions, the latter typically on upland plateaux.

The margins of ice sheets and ice caps, especially in Antarctica, frequently become afloat in the ocean that surrounds them. These floating sections, or ice shelves, normally cover shallow marine embayments or continental shelves with many islands. Eventually, slabs of ice break off from the ice cliffs that border ice shelves, thus producing (tabular) ICEBERGS.

The surface morphology and flow of glaciers (alpine glaciers, mountain glaciers) follow the morphology of the subglacial landscape. The morphology of ice fields, thinner than ice caps and lacking the characteristic ice dome, characteristically shows mountain uplands covered by ice, except for the highest ridges and peaks (see NUNATAK). Valley glaciers, elongated features confined by valley walls, occur as outlets of the ice-covered uplands and as glaciers on their own (Plate 51B). Their heads (highest section) often start in bedrock depressions (see CIRQUE) and their snouts (lowest section) reside within a valley. However, when the glacier emanates and terminates immediately beyond the valley mouth (i.e. beyond its lateral constraint), the ice tongue spreads outward to form a wide piedmont lobe, forming a piedmont glacier (Plate 51C). If outlet glaciers and valley glaciers extend offshore they form tidewater glaciers (Plate 51A). Cirque glaciers or glacierets are located within, or just barely extend beyond, the amphitheatre-like basins in which they form, and they tend to be wide rather than long (Plate 51D). Glaciers on steep mountain slopes above bedrock depressions, hanging glaciers or ice aprons, primarily loose mass as blocks of ice become detached and avalanche downslope. Sometimes, these and other avalanches supply the mass necessary for rejuvenated or regenerated glaciers to exist in locations where precipitation alone would be insufficient to support a glacier (Hambrey 1994; Sharp 1988; Sugden and John 1976).

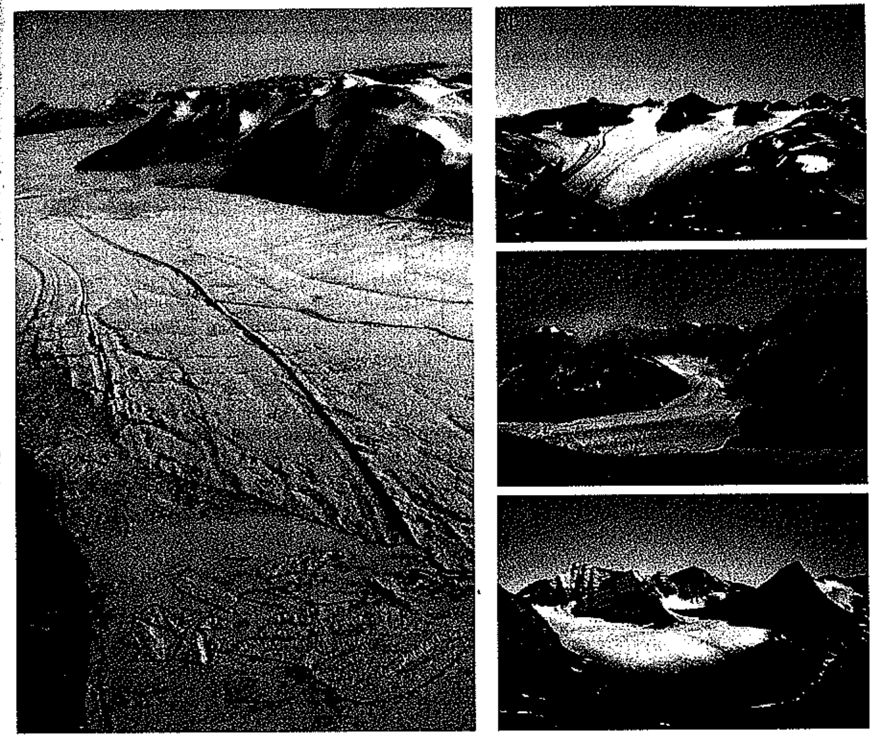


Plate 51 (a) outlet glacier, Antarctica; (b) valley glacier, Sweden; (c) piedmont glacier, Antarctica; and (d) cirque glacier, Sweden

Ordering glaciers by their thermal characteristics yields the (high) polar and temperate glacier-type end members, and the subpolar glacier type for those that combine both former characteristics. The ice in polar glaciers is below freezing throughout and meltwater is absent, except, maybe, for surface melting during short periods in summer. The ice in temperate glaciers is close to melting throughout, except, usually, for surface freezing during winter. Hence, the glacier has a temperature close to its melting point, which is 0 °C for pure ice at atmospheric pressure, but occurs at lower temperatures (pressure melting) when pressurized at depth. However, many glaciers experience both freezing and melting conditions; they are subpolar or polythermal glaciers. In polar regions they experience abundant surface melting and heating throughout the summer and in temperate regions they have a thick cold surface layer that is maintained throughout the summer (Ahlmann 1935).

Mass balance of glaciers

Most glaciers are nourished by snowfall and depleted by snow and ice melt. Because the conditions favourable for snow to fall and snow and ice to melt are related to the temperature of the atmosphere over the glacier surface, and because the mean annual temperature decreases with increasing altitude (lapse rate), the surface of a glacier normally shows two or more predictable zones

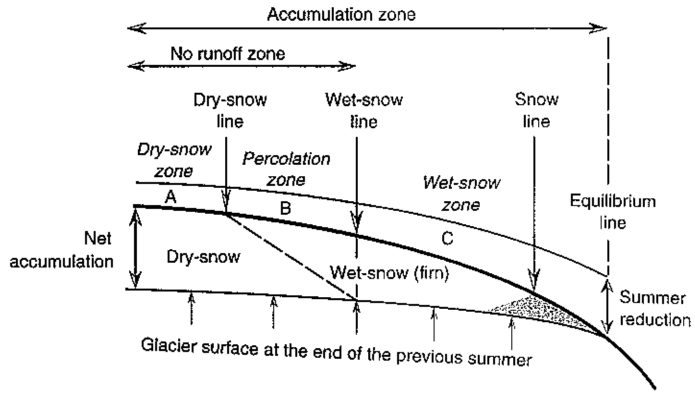


Figure 72 The surface structure of the accumulation area. The heavy line is the glacier surface at the end of the summer, the light line is the previous winter surface. The grey-shaded region is superimposed ice. Volume reduction by (A) recrystallization, (B) melt (and refreezing at depth) and (C) melt and runoff. Modified from Paterson (1994: 10)

(Figure 72). In the highest elevation zone, the accumulation zone, temperatures are at a minimum yielding conditions where the snowpack that accumulates during the winter or wet season is not melted entirely during the summer or dry season, thus creating an annual mass gain (accumulation). In the lowest elevation zone, the ablation zone, temperatures are higher and last winter's snowpack plus an additional amount of the underlying solid ice melts, creating an annual mass loss (ablation). The MASS BALANCE OF GLACIERS denotes the balance between the amount of snow remaining after the summer integrated across the accumulation area (and converted to the amount of water it represents when melted, m w.e. or metre water equivalent) and the amount of solid ice lost underneath the snowpack integrated across the ablation area (in m w.e.). When mass balance is positive, and more and more mass is added each year, the glacier will grow and expand. Conversely, when mass balance is negative over many years, the glacier will shrink and contract (retreat). When ice reaches considerable thickness a positive feedback mechanism occurs – because of the higher elevations the conditions for ice accumulation improve, especially close to the snout. The boundary between the accumulation and ablation zones is a relatively narrow zone, or line, where the annual mass balance is zero, the EQUILIBRIUM LINE OF GLACIERS (which exists on all glaciers that are not strongly out of equilibrium with contemporary climate, i.e. where the whole glacier surface becomes an accumulation area or ablation area). Some glaciers have additional boundaries, lines, in their accumulation areas (Figure 72). The dry-snow line borders the dry-snow zone, a region of extreme climate (in polar areas and at extremely high elevations), where no surface melt occurs, even in summer. The wet-snow line is the lower limit of the percolation zone, where snow melts at the surface and refreezes at depth during the summer. The refreezing of meltwater occurs at depth because the snowpack is initially cold, but the refreezing process releases heat and warms the snowpack until it reaches melting temperatures throughout by the end of the summer at the wet-snow line. Strictly taken, above the wet-snow line there is no mass loss (no runoff zone). The snow line is the lower boundary of the wet-snow zone, where all remaining snow at the end of the ablation season is at the melting temperature. The firn line, not shown, is the boundary between ice and firn at the end of the summer, and may coincide with the snow line (only in temperate regions where snow may transform to firn in one summer can this latter situation occur). Although there is mass

loss throughout the wet-snow zone during the summer, it is characterized by a positive annual mass balance. Meltwater that has refrozen at depth, forming ice lenses, may become particularly extensive and form superimposed ice layers underneath the firn in the wet-snow zone. These can visually crop out at the surface of a glacier at the end of the summer season between the snow line and the equilibrium line (Oerlemans 2001; Paterson 1994).

Alternative important components in the annual mass balance of glaciers are mass gain by avalanching and rime, and mass loss through avalanching, calving and sublimation.

Glacier flow

Structural glaciology, the geomorphology of glacier surfaces, shows that bodies of ice experience differential movements between the glacier bed and the surface, between the lateral margins and the glacier centre, and between different locations along its longitudinal profile. The most conspicuous features are crevasses, foliation structures, and band- or wave ogives (Forbes bands) (Hambrey 1994: 61–69; Paterson 1994: 173–190; Sugden and John 1976: 71–78). For example, a visible sign of glacier flow, except for these ice surface structures, is the creation of a BERGSCHRUND, an opening between (stagnant ice on) the valley wall and the glacier that pulls away from the wall as it moves downslope (Figure 73).

Crevasses, which are fissures in the ice surface with a typical depth of 25–30 m, occur abundantly on glacier surfaces, are often consistent in their direction over limited distances but vary considerably along the length of a glacier, and

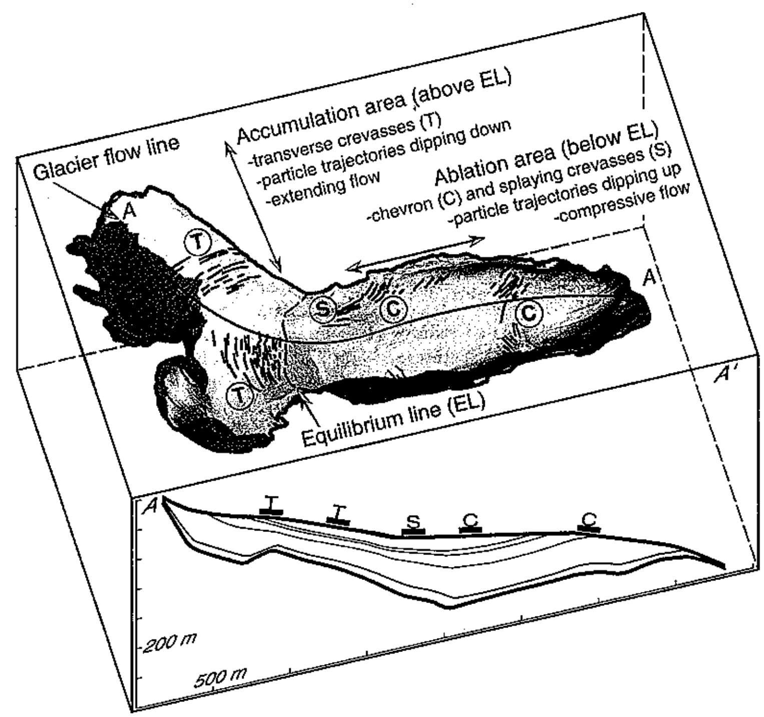


Figure 73 Glacier flow from surface structure and particle trajectories. Example is from Storglaciären, northern Sweden. Particle trajectories modified from Pohjola (1996). Note the location of the bergschrund (dashed) in the upper part of the accumulation area

will form where at least one principle stress is tensile and exceeds the tensile strength of the ice. The direction of crevasses can be predicted from a straightforward geometric analysis of stress distributions. Hence, the pattern of crevasses found in many accumulation areas of glaciers, transverse crevasses, can be related to the existence of extending flow (when ice flow accelerates downslope; i.e. where the glacier bed steepens, where the glacier is joined by another branch, on the outside of a glacier bend, and at the grounding line, where glaciers become afloat). Conversely, most glacier ablation areas reveal a pattern of crevasses, splaying crevasses, which are expected when the stress situation is dominantly compressive (when ice flow decelerates downslope; i.e. where the glacier bed flattens or rises, where the valley widens, on the inside of a bend, and at the snout; Hambrey 1994: 56). Finally, where the stress situation is neither dominantly extending, nor compressive, the lateral drag of the valley walls results in a third predictable stress and crevasse pattern, chevron crevasses, a feature common to middle sections of alpine glaciers. The width and depth of the crevasses depend primarily on the temperature of the ice, i.e. they are widest and deepest in polar glaciers.

Foliation structures result from previous structures that have been compressed (pure shear) and stretched (simple shear) to attain a direction that dominantly is parallel to ice flow. The original structures include the near-horizontal layering during accumulation (and the formation of ice lenses) and vertical or inclined elongated structures, such as crevasses, and point structures (moulins). Foliation structures, therefore, are consistent with differential motion through the ice body.

Forbes bands, or ogives, can form as glacier ice flows through an ice fall. Ice falls are causally related to the occurrence of very steep glacier beds. Here, the tensional stresses result in transverse crevassing and thinning while the ice moves through the fall. However, summer and winter conditions result in a modification of the ice as it moves through. Ice ablation during the summer results in thinner ice (and dirty because of dust collection) arriving at the foot of the fall. Precipitation during the winter, on the other hand, fills the crevasses with snow, resulting in thicker and cleaner ice arriving at the foot of the fall. For ice falls where the ice moves through within a year's time span, the resulting situation at the foot of the fall, where flow is compressive and crevasses are closed, are alternating ridges of blue ice (winter) and depressions of white ice (summer). These structures are convex in the direction of ice flow (even though the crevasses were slightly concave), indicating the differential motion in the transverse direction and, because of the crevassing, in the longitudinal direction.

The geomorphology of glacier surfaces, therefore, was the initial guidance in an understanding of some of the basic characteristics of glacier motion by the middle of the century (e.g. Tyndall 1860). Modern advancements in understanding the flow of glaciers, dating from the past fifty years, were mainly theoretical in nature (Glen 1952; Nye 1952; Weertman 1957), realizing that ice behaves like a crystalline solid, leading to theorems that have subsequently been verified experimentally (e.g. Raymond 1971).

From basic mass balance principles it must follow that a glacier of constant shape and volume (steady-state situation) must transfer all the annual mass gain in the accumulation area to the ablation area to compensate for the annual mass loss. For each transverse cross section through the glacier it is pertinent that the discharge of ice through that section per year balances the integrated annual mass balance across the up-glacier surface. Hence, this requires the discharge to be at a maximum at the equilibrium line, and normally this equates to having the thickest ice and highest ice flow velocities there, and implies the existence of extending flow in the accumulation area and compressive flow in the ablation area. Snow that falls in the accumulation area is successively compressed by subsequent layers of snow (until glacier ice is formed, which is incompressible), yielding a small velocity component towards the glacier bed. A cube of solid ice, when subjected to the range of pressures by the burden of overlying snow, firn and ice (typically less than 1 bar), will deform (stretching), thus departing on a trajectory that parallels, but is dipping slightly away from, the glacier surface (Figure 73). To satisfy mass continuity, this yields a convergence of flow in the accumulation area, and, typically, a thickening of the glacier downslope. Conversely, a cube of ice in the ablation area follows a trajectory which parallels, but is slightly directed towards, the slope of the ablating ice surface (Figure 73). This results in a divergence of flow in the ablation area, and, typically, a thinning of the glacier towards its lateral margin and snout.

As mentioned, the movement of ice is a gravity-driven internal deformation of the ice body. Because the deformation is at a maximum where the pressures are at a maximum, the deformation occurs primarily close to the glacier bed ('dragging' the overlying ice along). Because the amount of deformation higher up in the ice body, although dramatically less than at its bed, occurs in addition to the deformation for deeper ice layers, the velocity vector of ice flow close to the surface is measurably higher than at its bed.

Like other solids, ice deforms more readily when it is close to its PRESSURE MELTING POINT, a trait of temperate glaciers. Typical for pressure melting conditions is the presence of water (fluid state) and ice (solid), side by side, in the glacier body. The presence of water at the ice–bedrock interface is of importance in the motion of glaciers because its lubricating effect (especially when pressurized) facilitates the sliding of ice over its substrate (basal sliding). When ice covers sediment, basal sliding will normally be insignificant, but enhanced deformation of the water-saturated sediment may occur. Basal sliding would be very effective over a smooth bedrock surface, but, often, there are bedrock protrusions which hamper the effectiveness of the sliding process. When the obstructions are relatively small ($<10^{-2}$–10^{-1} m-size), ice melts on the upstream pressurized side, flows around the obstacle as water, and refreezes in its wake as the pressure drops, thus producing heat that can be used to help melting the ice on the upstream side. This experimental verification of the presence of regelation ice in the wake of obstacles is one of the modern findings of glacier flow. When the obstructions are relatively large (10^{-1}–10^{1} m-size), they create pressures so large that ice will deform more readily around it (by enhanced plastic deformation), a process that is less effective than regelation and thus more hampering to ice flow.

As noted, the effectiveness of glacier motion is in large part dependent on the temperature of the ice. Because most of the motion occurs in the basal ice as deformation and between the ice and the bedrock by sliding, it is specifically the subglacial temperature which is of interest. Temperate glaciers are warm-based (wet-), and their surface velocities integrate basal sliding and effective ice deformation. Polar glaciers are cold-based (dry-), which inhibits basal sliding and their surface velocities only reflect the integrated effect of ice deformation (at sub-optimal temperatures). Subpolar glaciers may have a bed at the pressure melting point and behave like temperate glaciers, or otherwise are cold-based and behave like polar glaciers.

References

Ahlmann, H.W. (1935) Contribution to the physics of glaciers, *Geographical Journal* 86, 97–113.

Glen, J.W. (1952) Experiments on the deformation of ice, *Journal of Glaciology* 2, 111–114.

Hambrey, M.J. (1994) *Glacial Environments*, London: UCL Press.

Nye, J.F. (1952) The mechanics of glacier flow, *Journal of Glaciology* 2, 82–93.

Oerlemans, J. (2001) *Glaciers and Climate Change*, Lisse: A.A. Balkema.

Paterson, W.S.B. (1994) *The Physics of Glaciers*, Oxford: Pergamon Press.

Pohjola, V.A. (1996) Simulation of particle paths and deformation of ice structures along a flow-line on Storglaciären, Sweden, *Geografiska Annaler* 78A, 181–192.

Raymond, C.F. (1971) Flow in a transverse section of Athabasca glacier, Alberta, Canada, *Journal of Glaciology* 10(58), 55–84.

Sharp, R.P. (1988) *Living Ice: Understanding Glaciers and Glaciation*, Cambridge: Cambridge University Press.

Sugden, D.E. and John, B.S. (1976) *Glaciers and Landscape: A Geomorphological Approach*, London: Edward Arnold.

Tyndall, J. (1860) *Glaciers of the Alps*, London: John Murray.

Weertman, J. (1957) On the sliding of glaciers, *Journal of Glaciology* 3, 33–38.

Further reading

Lliboutry, L. (1964–1965) *Traité de Glaciologie* (2 vols), Paris: Masson.

SEE ALSO: glacial deposition; glacial erosion; moraine

ARJEN P. STROEVEN

GLACIFLUVIAL (GLACIOFLUVIAL)

Glacifluvial is an adjective that applies to the processes, sediments and landforms produced by water flowing on, in and/or under glaciers and away from glacier snouts. Consequently, glacifluvial environments may occur in supraglacial, englacial, subglacial, ice-marginal and proglacial locations of alpine and continental glaciers both present and past. Proglacial, glacifluvial environments are often transitional to fluvial environments when the processes, sediments and landforms become dominated by non-glacial tributary inflows rather than the annual rhythm of melting ice (Lundqvist 1985).

Meltwater supply and pathways

Water enters a glacier from melting ice, snow melt, rainfall, hillslope runoff and the release of stored water. Consequently, most glacifluvial environments are dominated by strong seasonal, diurnal and episodic discharge variations. The release of stored water from subglacial, englacial, supraglacial or ice-marginal lakes and reservoirs produces floods several orders of magnitude greater than 'normal' melt-related flows. Such floods and megafloods (peak discharges estimated at 10^6–$10^7\,m^3.s^{-1}$ compared to $10^5\,m^3.s^{-1}$ for the Amazon today) are known as OUTBURST FLOODS or jökulhlaups.

The path of meltwater through a glacier is determined by water supply (amount and location), ice temperature and dynamics and basal substrate. Pathways include: (1) supraglacial meandering channels; (2) englacial passages; (3) subglacial channels cut up into the ice (ice tunnels) or down into the bed, broad flows (sheets), linked-cavities, films, canals or aquifers; and (4) proglacial channels, broad flows and jets (into standing water – lakes, oceans).

Glacifluvial processes

Glacifluvial processes include erosion, transport and deposition. The type, rate and effectiveness of meltwater erosion are influenced by the nature of the basal substrate (sediment, bedrock), meltwater supply and pathway, and sediment supply. Mechanical erosion is effective because rapidly flowing meltwater in channels and broad flows is turbulent and carries a high sediment load (see SEDIMENT LOAD AND YIELD). Mechanical erosion includes: (1) hydraulic action – the force of water against its bed lifts or drags loose debris into the flow; (2) CAVITATION – shock waves and microjets resulting from the formation and implosion of bubbles of vapour (cavities) cause rock pitting or loosen mineral grains; and (3) abrasion – the impact and grinding of rock particles carried by the flow on themselves (attrition) and on flow boundary materials causes wear. Chemical erosion, or DISSOLUTION, occurs when meltwater removes soluble minerals in bedrock and debris. Subglacial dissolution is particularly effective because freshly abraded bedrock and debris present a high surface area for chemical reactions and solutes are continually flushed from the system. Together, meltwater erosion processes act to fracture, round and wear down bedrock and rock fragments and result in a variety of erosional landforms.

Landforms resulting from meltwater erosion

Meltwater erosion produces meltwater channels, bedrock erosion marks (s-forms) and may form some DRUMLINS, flutings, ribbed and hummocky terrain (Shaw 1996).

Meltwater channels may form (1) along the ice margin, (2) proglacially, or (3) subglacially. These channels exist in a variety of sizes (cm to km wide, cm to hundreds of m deep, decimetres to tens of km long) and substrates (bedrock, sediment); channels associated with past ice sheets are typically larger than those associated with alpine glaciers today. They may be differentiated by their slope and relationship to other glacial landforms. Ice-marginal (lateral) channels typically form parallel to the ice margin of cold-based glaciers, are often left perched and nested on valley sides during glacier retreat and their slope approximates that of the glacier surface at the time of formation. Advancing glaciers may divert rivers parallel to their ice front forming URSTROMTÄLER. Proglacial channels can form braided patterns (multiple shallow channels separated by bars) across outwash plains and always follow downslope paths (Plate 52a). Catastrophic drainage of ice-dammed or proglacial lakes has produced (1) trench-like valleys or spillways, and (2) networks of dry channels and waterfalls (see SCABLAND; Plate 52b). Subglacial channels may follow upslope paths, cross-cut drumlins and contain ESKERS (Plate 52c). TUNNEL VALLEYS or channels are long (can be >100 km), wide (<4 km), flat-bottomed, overdeepened (<100 m), often radial or anabranched subglacial channel systems that formed under ancient ice sheets. They can be buried by thick fills, muting their topographic expression. They may have formed catastrophically by meltwater erosion during the waning, channelized stages of subglacial megafloods (Shaw 1996).

Bedrock erosion marks, or s-forms, come in a variety of shapes and sizes (cm to km scale) and are found on the beds of bedrock channels and on broad bedrock plains. They are classified according to their shape and to the direction of formative

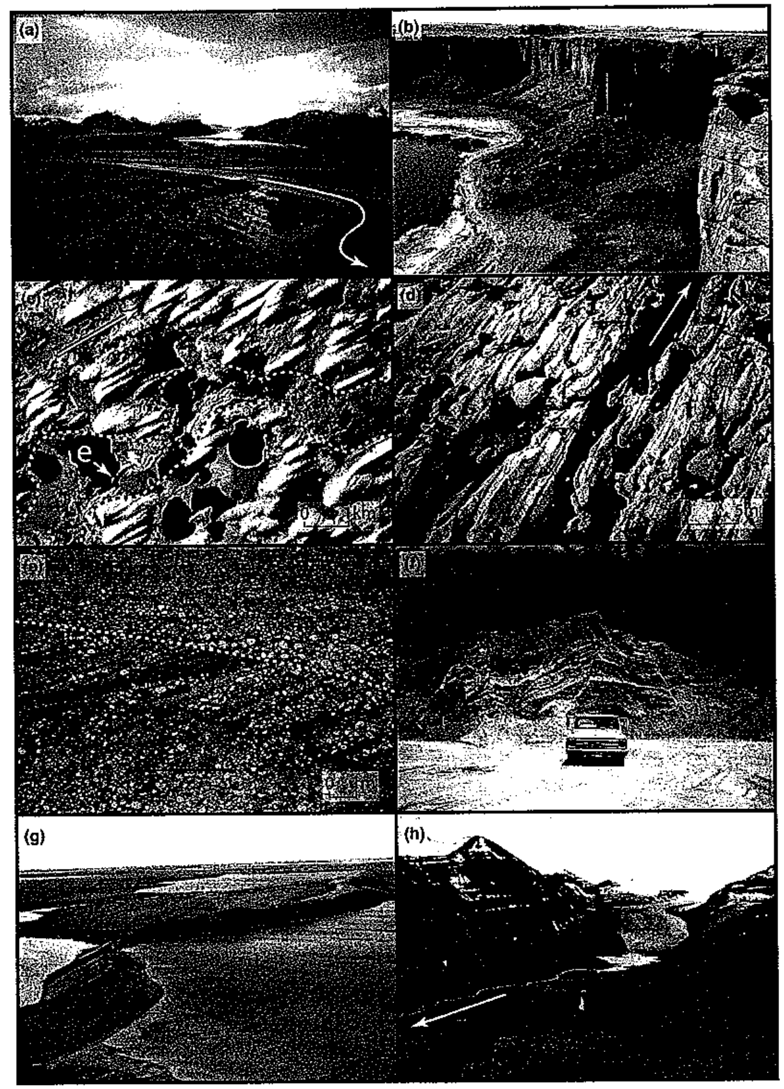


Plate 52 (a) braided outwash on Icelandic sandur (Skeiđarársandur, courtesy C. Simpson); (b) dry falls along Shonkin Sag spillway (USA, backwall ~55 m high); (c) cavity-fill drumlins truncated by tunnel channel (dashes) containing esker [e] (Livingstone Lake, Canada; aerial photograph A14509-77 © Her Majesty the Queen in Right of Canada, Centre for Topographic Information (Ottawa), Natural Resources Canada); (d) bedrock erosion marks (French River, Canada; Shaw 1996, © Wiley and Sons Ltd. Reproduced with permission); (e) cross-stratified [x] and horizontally-bedded [h] cobbles (Harricana esker, Canada); (f) faulted sand and gravel lithofacies (Campbellford esker, Canada); (g) sinuous esker (Victoria Island, Canada); (h) valley train (Saskatchewan glacier, Canada). Arrows indicate direction of formative water flow

flow. Mussel shell-shaped scours (muschelbrüche), sickle-shaped scours (sichelwannen; Plate 52d), comma-shaped scours and transverse troughs form mainly transverse to flow direction. Rock drumlins, rat-tails (tapering rock ridges extending from resistant rock knobs), flutes (spindle-shaped scours), furrows (Plate 52d) and cavettos (channels on vertical walls) form parallel to flow. POT-HOLES record vertical scour. Together s-forms often exhibit a directional consistency, and a hierachical and systematic arrangement consistent with erosion by turbulent subglacial megafloods (Shaw 1996; Figure 74).

Drumlins are elongated hills of various sizes (up to ~2 km long, tens of m high, hundreds of m wide) and shapes (inverted spoon, parabolic, transverse asymmetrical, spindle-shaped; Plate 52c). They occur in en-echelon patterns and in fields spanning hundreds of kilometres. Some flutings are long (tens of km), narrow (hundreds of m) remnant ridges that occur downflow from escarpments. Ribbed terrain is composed of fields of coalescent and subparallel, convex-upflow and crenulate-downflow ridges (up to 30 m high) that are formed transverse to flow. Hummocky terrain is identified as a field of mounds (<10 m high and ~100 m diameter) and hollows. These subglacial bedforms are often transitional to one another and the material within them may be truncated at the landsurface. Consequently, some drumlins (inverted spoon-shaped drumlins), flutings, ribbed and hummocky terrain are attributed to erosion by turbulent subglacial megafloods (Shaw 1996; Figure 74).

Glacifluvial sediment

Glacifluvial sediment is mainly derived from (1) material supplied to the glacier surface from valley side rock falls, debris flows and avalanches, or (2) meltwater erosion of bedrock, sediment or debris-rich ice along its flow path. Typically, sediment transport varies with glacial environment and discharge, being greatest in subglacial channel and broad flows and during OUTBURST FLOODS. Deposition rates are generally greatest at the ice margin where meltwater issues from ice tunnels onto open outwash plains or into standing-water bodies.

Glacifluvial sediment is deposited in BEDFORMS which vary in scale from centimetres (e.g. ripples) to hundreds of metres (e.g. bars or macroforms). Bedforms are preserved in the sedimentary record as lithofacies (Plate 52e). Lithofacies are sedimentary units distinguished by their physical and/or chemical characteristics such as colour, texture, structure and mineralogy. By applying experimental relationships derived from pipe and flume studies, lithofacies are used to reconstruct former bedforms, flow conditions (see PALAEOHYDROLOGY) and glacifluvial environments. For example, cross-laminated sand records ripples deposited during low flow, whereas horizontally bedded gravel may record the movement of gravel sheets across the tops of macroforms during flood flows (Plate 52c).

Glacifluvial sediment may be found in ice-contact environments (supraglacial, englacial, subglacial and ice-marginal) or proglacial environments beyond an apron of stagnant ice. It shows many of the same characteristics as sediment of non-glacial rivers – both are typically characterized by sorted and stratified sand and gravel lithofacies. However, glacifluvial deposits differ from non-glacial fluvial deposits in several ways. (1) Glacifluvial sediment is typically coarse grained (boulder through sand sizes) as the flow velocity is generally too high for settling of the finest particles (silt and clay); such particles become trapped in lateral and distal standing-water bodies. (2) Course-grained lithofacies may exhibit relatively poor sorting (large grain-size range) and rudimentary bedding when rapidly deposited during the waning stages of an outburst flood. (3) Glacifluvial sediments typically exhibit abrupt changes in lithofacies (Plate 52e) due to pronounced seasonal and episodic changes in flow regime. (4) Ice-contact glacifluvial deposits frequently include flow deposits (diamicton) and till balls (eroded ice deposits) and exhibit structures indicative of shearing, faulting (Plate 52f), slumping and subsidence. These characteristics develop due to the proximity to a moving glacier and to the melting of buried or supporting ice.

Landforms resulting from meltwater deposition

Meltwater deposition forms ice-contact and proglacial landforms. Ice-contact landforms may include drumlins, ribbed and hummocky terrain, some crevasse-fill ridges and De Geer MORAINES, eskers, KAMES, kame terraces, outwash fans and some end, grounding-line and interlobate moraines. Proglacial landforms include outwash plains, valley trains and sandurs. The main differences between these two categories are (1) ice-contact landforms may contain structures

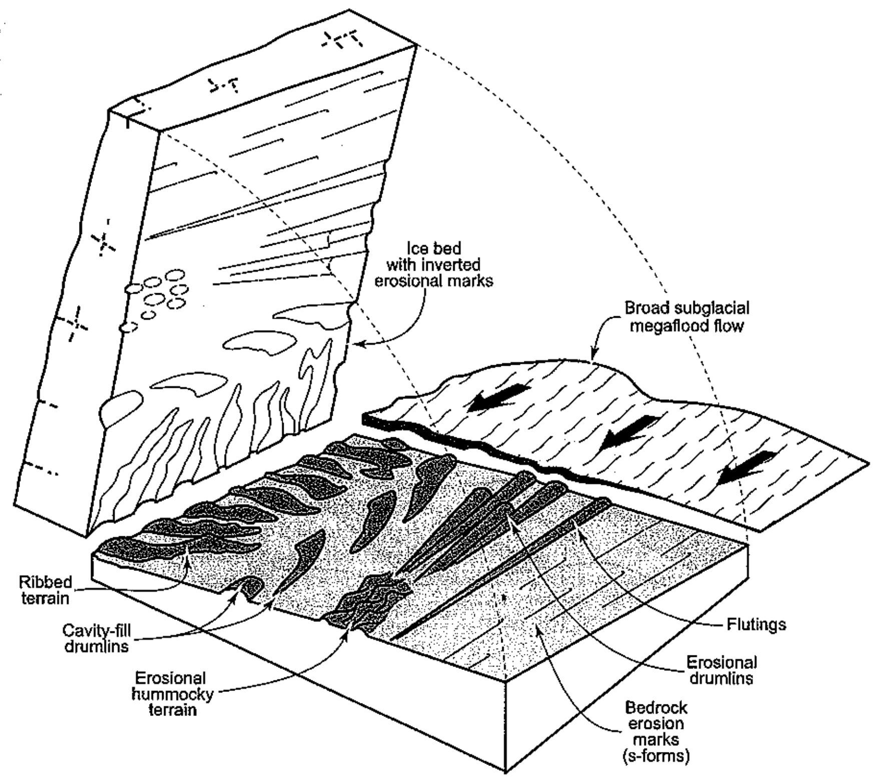


Figure 74 A model for subglacial landforms produced by broad subglacial megafloods (modified from Shaw 1996)

indicative of proximity to a moving glacier (e.g. thrust faults and shears) and/or the melting or removal of buried or supporting ice (faulting, slumping, pitted surfaces; see KETTLE AND KETTLE HOLES), and (2) proglacial landforms may contain sediment indicative of both glacifluvial and non-glacial processes.

Some drumlins (spindle, transverse asymmetrical and parabolic forms; Plate 52c), ribbed and hummocky terrain may have formed by deposition into cavities incised into the ice base during subglacial megafloods (Shaw 1996). As the broad megafloods waned, sediment carried in the flow was rapidly deposited into flow parallel, transverse and non-directional cavities forming cavity-fill drumlins, ribbed and hummocky terrain respectively (Figure 74). These landforms are composed of glacifluvial sediment with a sedimentary architecture that conforms to the landform shape.

Crevasse-fill ridges are linear, low ridges (up to ~10 m high) that are arranged in a pattern that mimics the radial and transverse crevasse patterns of a glacier. They formed by the infilling of crevasses within or at the base of glaciers (Figure 75). De Geer moraines are linear, low (<10 m) ridges that occur in fields subparallel to one another and in locations where the glacier was in contact with a standing-water body. They may form at the glacier margin during punctuated glacier retreat or in subglacial crevasses. Both ridge types may contain glacifluvial sediment.

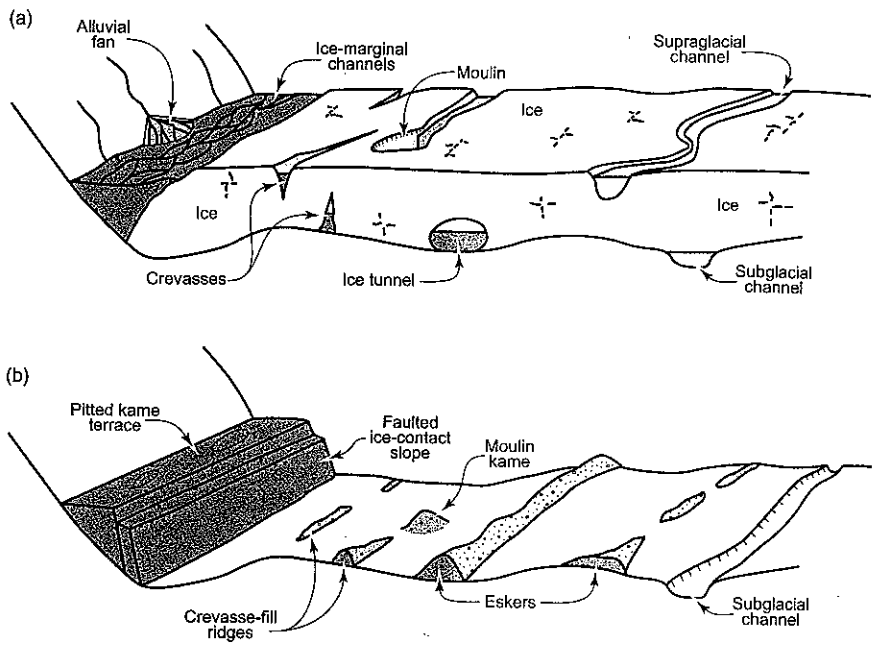


Figure 75 Development of glacifluvial ice-contact landforms (a) before and (b) after stagnant ice melt

Eskers form narrow, sinuous ridges (Plate 52g) or a series of ridges separated by broader beads. They occur in a range of sizes (m to tens of m high, m to hundreds of m wide, tens of m to hundreds of km long). They can be located in valleys or follow upslope paths. They may occur in isolation or in groups forming subparallel, deranged (not aligned with regional ice flow) or dendritic (tree-like) patterns. Eskers record the location of past ice-walled streams – mainly subglacial streams as the ridge-form is often lost when supraglacial or englacial channel deposits are lowered to the ground during glacier melt (Figure 74). Narrow ridges are tunnel deposits, whereas beads contain macroform, fan or delta sediments that formed where tunnel flow expanded into a subglacial cavity or at the ice margin. Strings of beads may indicate the punctuated retreat of the glacier front. Deranged eskers likely formed under stagnant ice. Long (hundreds of km), dendritic esker systems may have (1) developed in long and persistent (perhaps operating for hundreds of years) subglacial tunnels in stagnant ice or (2) formed almost instantaneously (over perhaps weeks or months) during the drainage of large subglacial reservoirs or supraglacial lakes.

Kames are steep-sided, variously shaped mounds of sand and gravel that were originally deposited with two or more ice-contact margins (Figure 75). Examples include moulin deposits (moulin kames) and small deltas or fans deposited at or on the ice margin (delta kames). Kame terraces are linear, often pitted benches of sand and gravel deposited by braided rivers which flowed between the valley side and the ice margin. They need not be altitudinally matched on both sides of a valley glacier.

Outwash fans are fan-shaped bodies of downstream-fining sediment with their apex at a meltwater portal. Deposition on land results in subaerial outwash fans and deposition in water (at a grounding line) results in subaqueous outwash fans. Adjacent outwash fans may coalesce forming a ridge along the ice front – an end moraine (on land) or grounding-line moraine (in water), or between ice lobes – an interlobate moraine. Past ice sheets have left many such glacifluvial moraines. Subaerial outwash fans often grade downflow into proglacial stream deposits.

Outwash plains are planar landforms containing proglacial stream deposits. Proglacial streams are often braided (Plate 52a, h) as high sediment loads, fluctuating discharge and a lack of vegetative anchoring results in a high degree of channel instability. Channels vary in width (m to hundreds of m) and bars vary in size (m to hundreds of m long, m relief). Braided streams continually evolve with each successive flood by channel scour and fill, bar development and overbank deposition. Where numerous proglacial streams issue from the ice front onto an open lowland an extensive outwash plain known as a sandur is formed (Plate 52a). When proglacial streams are hemmed in by valley sides in mountainous terrain, deposition is focused along the valley axis resulting in thick valley fills and a linear outwash plain called a valley train (Plate 52h). Where proglacial rivers enter standing-water bodies deltas (see GLACIDELTAIC) may form.

Outwash plains typically exhibit downflow changes in their morphology and composition reflecting a lessening of glacier influence and a decrease in energy away from the glacier snout. Close to the glacier (proximal) coarse gravel devoid of vegetation is arranged into longitudinal bars separated by a few large channels. The surface is often pitted. Grain size is highly variable reflecting strong melt and flood cycles. With increasing distance from the ice front transverse bars separated by a complex network of braided channels become prevalent, grain size decreases (sand and gravel are present) and grain roundness increases due to selective sorting and abrasion during transport. Lithofacies variation within and between beds is diminished as inflowing non-glacial tributaries dampen discharge fluctuations. Distally (furthest from source), flow in shallow braided channels, sheets or meandering streams deposits sand.

Megafloods and climate change

A broad suite of glacial landforms are now attributed to megafloods from past ice sheets (Figure 74; other explanations are also debated, see DRUMLINS; SUBGLACIAL GEOMORPHOLOGY). The discharge of

such enormous quantities of freshwater across continents and into the oceans caused sea level to rise and may have modified ocean and atmospheric circulation, heralding climate change. As meltwater drainage, ice temperature and dynamics (movement) are linked, glacifluvial sediments and landforms contain a record of ice and water behaviour that is essential in our quest to understand climate change.

References

Lundqvist, J. (1985) What should be called glaciofluvium? *Striae* 22, 5–8.
Shaw, J. (1996) A meltwater model of Laurentide subglacial landscapes, in S.B. McCann and D.C. Ford (eds) *Geomorphology Sans Frontières*, 183–226, Chichester: Wiley.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.
Drewry, D. (1986) *Glacial Geologic Processes*, London: Arnold.

SEE ALSO: esker; drumlin; glacier; kame; meltwater and meltwater channel; outburst flood; scabland; subglacial geomorphology; tunnel valley

TRACY A. BRENNAND

GLACILACUSTRINE (GLACIOLACUSTRINE)

Modern and ancient glacilacustrine deposits tend to be variable in grain size, mineralogy, bedding thickness and sedimentary structures, reflecting the broad range of settings in which they accumulate. Glacial lakes may originate from ice erosion of bedrock, in depressions of glacial deposits, or be impounded behind drainage barriers composed of moraine, outwash or ice (Hutchinson 1957). Today, lakes of glacial origin outnumber all other lake types combined. However, most present-day glacial lakes owe their origin to Pleistocene glacial activities and are now under no direct influence of glaciers. Thus it is useful to distinguish glacial lakes and their deposits from those of glacier-fed lakes, and to divide the latter into those bordered by an actively calving glacier (ice-contact or proglacial lakes) and those located downstream (non-contact or distal lake) (Ashley *et al.* 1985).

Most of the material deposited in glacial lakes comes from sediment in suspension and bedload in glacial meltwater streams. Additional contributions may be derived from slope processes delivering sediment directly into the lake (slope wash, avalanching, debris torrents, for example), atmospheric precipitation (including volcanic events), hydrochemical precipitation, biogenic activity, upwelling of material from groundwater flow, and resuspension from bottom current activity.

Deltas form where a meltwater stream or the glacier itself enters a lake. Sudden flow expansion causes an abrupt decrease in stream velocity and competence, which in turn results in rapid deposition of coarser material (see GLACIDELTAIC (GLACIODELTAIC)). At ice margins, other glacilacustrine sediments are also deposited, including subaquatic flow tills, formed by gravity deposits from debris-rich glacier ice standing in a lake. Icebergs can release particles either individually, dropstones, or in conical debris mounds on the lake floor.

The bulk of sediment discharge into a glacial lake comes from glacial streams during the spring and summer-melt period. Concentrations of suspended sediments are highly variable, with values ranging from a few $\mathrm{mg\,l^{-1}}$ to $\mathrm{g\,l^{-1}}$ in extreme cases. Density differences between inflowing stream waters and glacial lakes result largely from differences in suspended sediment concentrations and temperature. With strong density contrasts, the incoming stream water will maintain its integrity and flow into the lake as a discrete density current, either as an overflow (if its density is less than the lake water), an interflow (strong thermal stratification may result in flow along the thermocline), or underflow (if the inflowing water is more dense). The highly seasonal and weather-dependent nature of glacial-river discharge, temperature and suspended sediment concentration, together with the normal seasonal evolution of lake thermal structure, result in changing and often complex mixing and sedimentation patterns at different stages of the year. The resulting rhythmic deposition of sediments is a signature of many ice-contact and distal glacier-fed lakes.

Turbid underflows, high-density currents generated by underflowing sediment laden river water which produce quasi-continuous currents, and episodic surge-type currents formed by subaqueous slumping (velocities may range up $1\,\mathrm{m\,s^{-1}}$) both transfer suspended sediment and a large quantity of bedload directly to deeper parts of the lake floor. A distinctive suite of graded deposits often characterized by ripple-drift and cross-laminations result. In lakes where underflows dominate, the descent of turbidity currents down the basin sides may inhibit deposition and in places may cause active erosion.

When and where underflow activity is not evident, such as during winter months or due to fluctuations in discharge, settling of particles takes place from sediment suspended in the water column. The resulting deposits, normally only a few millimetres to centimetres thick, grade from silty-clay at the base to fine clay at the top. They often terminate abruptly with a sharp contact, due to a new underflow influx of coarse material. In the most distal areas of glacial lakes, variations in sediment inflow may be sufficiently damped to give rise to homogeneous clays.

A signature of many glacial lake floor deposits are 'rhythmites'. These are pairs (couplets), composed of light-coloured, silt layers, representing spring flood or storm deposits, and dark, clay layers, with higher organic content, representing quiet deposition under winter ice. The contact between the two layers may be gradational, but more often it is sharp. Multiple laminations may occur within the more proximal silt layers, reflecting short-term fluctuations (hours and days) in sediment influx and dispersal. Local factors, load and volume of the meltwater stream, the depth of the lake and relief of its floor, the strength of the currents and the distance from the point of entry into the lake, affect the thickness of the couplets (Menounos 2002). A recurring theme in discussions of rhythmites is their periodicity. De Geer (1912) introduced the term 'varves' to describe annual couplets. Non-annual glacilacustrine rhythmites can be formed from sudden fluctuations of discharge and sediment load, sometimes from OUTBURST FLOODS, cold and warm spells of a non-annual nature, episodic slope activity, or periodic action of storms stirring up lake waters (Sturm 1979). Great care must be taken to establish a reliable, independent chronology for rhythmites, especially if they are to be used as a geochronological tool (Brauer and Negendank 2002). Varved glacilacustrine deposits have been used to interpret high-resolution records of paleoenvironmental conditions; notably, climate, glacial activity, mineralogy of drainage areas, and changes in water level, temperature and trophic state (see, for example, Karlen 1976; Leonard 1986).

Shoreline processes in glacial lakes are similar to those in lakes in other environments. Lake waters standing at particular levels create strandlines with wave-eroded scarps, beaches, small deltas and terraces. Coarse-washed gravel, cobble and boulder deposits may accumulate where waves erode older glacigenic (e.g. till) deposits. In glacial lakes, wave activity may be inhibited for part of the year by the presence of ice cover. The effects of movement of ice cover against the shore, due either to thermal expansion or wind coupling, produce small ice-push features, which may reach heights up to a few metres. The inclination of glacial strandlines (commonly 1 or $2\,\mathrm{m\,km^{-1}}$) gives important insight into the rebound and tilting since ice unloaded certain areas.

Water levels in many ice-contact lakes fluctuate widely, a consequence of meltwater filling and subsequent ice-dam collapse and drainage. This has important effects on lake-bottom sediments, through scouring and slumping, as well as ancillary effects due to changing wave base, iceberg grounding and adjustments of distribution patterns of suspended sediments.

References

Ashley, G.M., Shaw, J. and Smith, N.D. (1985) Glacial sedimentary environments, Society of Paleontologists and Mineralogists, Short Course 15, Tulsa, OK.

Brauer, A. and Negendank, J.F.W. (2002) The value of annually laminated lake sediments in paleoenvironmental reconstruction, *Quaternary International* 88, 1–3.

De Geer, G. (1912) A geochronology of the last 12,000 years, 11th International Geological Congress (Stockholm, 1910) 1, 241–1, 258.

Hutchinson, G.E. (1957) *A Treatise on Limnology. Geography, Physics and Chemistry*, New York: Wiley.

Karlen, W. (1976) Lacustrine sediments and tree-limit variations as indicators of Holocene climatic fluctuations in Lappland, Northern Sweden, *Geografiska Annaler* 58A, 1–34.

Leonard, E. (1986) Varve studies at Hector Lakes, Alberta, Canada, and their relationship to glacial activity and sedimentation, *Quaternary Research* 25, 199–214.

Menounos, B. (2002) Climate, fine-sediment transport linkages, Coast Mountains, British Columbia, Ph.D. Thesis, Department of Geography, The University of British Columbia, Vancouver, Canada.

Sturm, M. (1979) Origin and composition of clastic varves, in C. Schlüchter (ed.) *Moraines and Varves*, 281–285, Rotterdam: Balkema.

SEE ALSO: glacier; glacideltaic; glacifluvial

CATHERINE SOUCH

GLACIMARINE (GLACIOMARINE)

Here, the term glacimarine is preferred to other alternatives (glacial marine, glacial-marine and glaciomarine) because etymologically, words with Latin roots are joined with an 'i'. An inclusive definition is also preferred here, where the term is taken to encompass the environment, processes and deposits including landforms, sedimentary systems, stratigraphy and life forms. Glacimarine systems include a combination of glacial and marine processes that produce a penecontemporaneous mixture of primarily siliciclastic and biogenic sedimentary deposits. Terrestrial sediment is introduced by ice rafting and rainout of debris (IRD); by fluvial transport feeding turbid overflow plumes with eventual suspension settling of particles; by mass flows and rockfalls from ice contact and shoreline subaerial systems; by aeolian transport with eventual settling through water (perhaps via sea ice); and by shoreline and shelf processes such as longshore transport. Glacimarine settings occur within a range of climatic (and glaciological) regimes from polar, to subpolar, to cool temperate, and encompass fjords and nearshore areas, continental shelves and the deep sea.

Grounding-line depositional systems are formed at the contact of a glacier with the seafloor. These deposits take the form of a bank (morainal bank (less-favoured alternatives: moraine, submarine moraine and moraine bank)), a fan (grounding-line fan (less-favoured alternatives: subwash fan, glacimarine fan and submarine ice-contact fan)) and a wedge (grounding-line or grounding-zone wedge and trough-mouth fan (less-favoured alternatives: till tongue, till delta, subglacial delta and diamict apron)). Grounding-line systems include a mixture of facies: till, glacimarine diamicton (stratified or massive), gravelly mud (laminated, e.g. cyclopels and cyclopsams, or massive), poorly or well-sorted sand and gravel (stratified or massive), and interlaminated sand and mud (e.g. turbidites) (see Further reading for details).

Till with various modifiers (e.g. waterlain till and paratill), has been used as the genetic term for glacimarine diamicts; however, till is best reserved for deposition directly from glaciers without modification such as by flowage or by currents during rainout. Thus glacimarine diamict is preferred, and if genetic interpretations are possible, then such terms as debris flow deposit (or debrite), or rainout diamict (produced from ice rafting), or ice-keel turbate (produced by keels of icebergs or sea ice) may be used. Specific environmental terms for the rainout diamicts may be shelfstone diamict or bergstone diamict, depending on whether their debris source is an ice shelf or icebergs, respectively.

Beyond these ice-contact systems that extend two to several kilometres from a grounding line, are ice-proximal (to ~10 km from a grounding line) and ice-distal glacimarine systems (to thousands of km from a grounding line, e.g. Heinrich layers). These distances are relative to grounding lines and may be within an ice shelf or an iceberg zone. The main glacial components are from IRD, suspension settling and, more rarely, wind transport. Deposits are either gravelly mud or diamict depending on relative accumulation rates of IRD and matrix sediment, which often is from meltwater streams. The matrix is stratified under higher current strengths and sedimentation rates or under continuous ice cover (which control the degree of bioturbation), and is otherwise massive. However, extremely high sedimentation rates with few bottom currents can produce massive deposits.

Ice rafting occurs via three forms of ice and, if possible, recognizing the distinction is useful, such as by: ice shelves and floating glacier-tongues – ISRD, icebergs – IBRD, and sea ice – SIRD. Sea-ice rafting perhaps should be excluded from glacimarine systems because it is not strictly glacial and may occur under non-glacial conditions. However, often distinguishing SIRD from other IRD is impossible and thus it is commonly included in the glacimarine system. Ensuring that particle rafting is not by tree roots or kelp hold-fasts is also important. A French term, *glaciel* has been suggested for sediment containing IBRD and SIRD, but is not commonly used.

Biogenic components in glacimarine deposits become more common with lower terrestrial sediment flux and meltwater; that is, either with distance from a glacier terminus or in colder climates. The geologically significant components include various macrofossils and microfossils, but diatoms commonly dominate and often form diatomaceous mudstone and diatomaceous ooze (diatomite). Marine productivity and diversity may depend on sea-ice extent, thickness and seasonal longevity, on sea water temperature and salinity changes, and on current up-welling (including polynya); thus these records contain high-resolution climate signals.

Morphologically significant forms produced in glacimarine settings include: fjords, cross-shelf troughs (or submarine troughs or sea valleys), inter-ice-stream ridges, mega-lineations (large-scale forms like flutes), flutes, grounding-line systems, iceberg and sea-ice scours, ploughs, or wallows, and striated boulder pavements.

The glacimarine environment includes sedimentary systems and processes that are typical of lower latitude settings, such as deltas, fan deltas, estuaries, tidal flats, linear sandy shorelines, shelves and deep water systems that commonly may include indicators of ice action described above. It includes lags, erosional surfaces, hiatuses and condensed sections produced from reworking by marine currents, from sediment starvation under large ice shelves or in ice distal areas during glacial retreat, and from isostatic rebound. By analogy with terrestrial glacial outwash and lacustrine systems, paraglacial marine settings occur where glaciers terminate on land, but their products of glacial rock flour accumulate as marine mud, perhaps including SIRD.

Further reading

Anderson, J.B. (1999) *Antarctic Marine Geology*, Cambridge: Cambridge University Press.
Anderson, J.B. and Ashley, G.M. (eds) (1991) *Glacial Marine Sedimentation: Paleoclimatic Significance*, Special Paper 261, Boulder, CO: Geological Society of America.
Davies, T.A., Bell., T., Cooper, A.K., Josenhans, H., Polyak, L., Solheim, A., Stoker, M.S. and Stravers, J.A. (eds) (1997) *Glaciated Continental Margins: An Atlas of Acoustic Images*, New York: Chapman and Hall.
Dowdeswell, J.A. and Ó Cofaigh, C. (eds) (2002) *Glacier Influenced Sedimentation on High Latitude Continental Margins*, Special Publication No. 203, London: Geological Society.
Dowdeswell, J.A. and Scourse, J.D. (eds) (1990) *Glacimarine Environments: Processes and Sediments*, Special Publication No. 53, London: Geological Society.
Molnia, B.F. (ed.) (1983) *Glacial-Marine Sedimentation*, New York: Plenum Press.
Powell, R.D. and Elverhøi, A. (eds) (1989) Modern Glacimarine Environments: Glacial and Marine Controls of Modern Lithofacies and Biofacies, *Marine Geology* 85, III-416.
Syvitski, J.P.M., Burrell, D.C. and Skei, J.M. (1987) *Fjords: Processes and Products*, Berlin: Springer-Verlag.

ROSS D. POWELL

GLACIPRESSURE (GLACIOPRESSURE)

The term, 'Glaciopressure' was introduced by Panizza (1973) to indicate the pressure of ice on the narrow part of a valley, which is particularly intense at the confluence of glacial tongues in the areas affected by Pleistocene glacier advances. It caused rock deformations in correspondence with surfaces of structural discontinuity, like strata, fissures, etc., favouring the formation of sliding surfaces. In fact, some landslides which took place in the late Glacial and Post Glacial were observed in the Alps, and particularly in the Dolomite region: they were triggered by a tensional discharge following the loss of pressure previously exerted on the rocky slopes by two or more glaciers merging in a valley narrow. Even if the fall of large slope portions can directly affect human settlements or obstruct a whole valley, with the negative resulting consequences, the extremely high risk degree assumed by this type of phenomenon is purely theoretical. Indeed, the long time span from the withdrawal of the glacial network to the present has practically produced the total exhaustion of these events.

References

Panizza, M. (1973) Glaciopressure implications in the production of landslides in the Dolomitic area, *Geologia Applicata à Idrogeologia* 8(1), 28–298.

MARIO PANIZZA

GLACIS D'ÉROSION

Glacis d'érosion are a form of PEDIMENT, a gently inclined slope of transportation and/or erosion that truncates rock and connects eroding slopes or scarps to areas of sediment deposition at lower levels (Oberlander 1989). Two fundamental types of pediment are recognized by Oberlander (1989): glacis d'érosion, which truncate softer rocks adjacent to a more resistant upland; and 'true' pediments, where there is no change in lithology between upland and pediment.

The name glacis d'érosion is derived from the work of French geomorphologists who studied examples of these landforms on the northern margin of the Sahara Desert, where they are particularly well developed on the flanks of the Atlas Mountains (Coque 1960). These landforms truncate weak materials such as poorly indurated Tertiary sediments, and tend to be veneered by alluvial gravels, indicating the importance of fluvial processes in their creation (Dresch 1957). The glacis piedmonts of the Atlas Mountains have a distinctive morphology, consisting of a series of coalescing flattened cones whose apices occur where stream channels debouch from upland drainage basins. The glacis long profiles range from nearly rectilinear to concave; the latter form having a slope of about 10 degrees at the top, dropping to about 3 degrees or less at the base.

Glacis d'érosion often exhibit multiple levels, or terraces, which can be traced back into the upland drainage basin where they form river terraces (see TERRACE, RIVER) (Plakht *et al.* 2000). These forms, known as stepped or nested glacis (Coque and Jauzein 1967), are formed as older glacis are incised by stream channels, which then form a younger glacis at a lower level, the new glacis being inset within the older glacis. The resulting landform appears similar in form to a telescopically segmented ALLUVIAL FAN, leading some workers to suggest that both landforms result from similar responses to environmental change (White 1991). The slope profiles of stepped glacis tend to converge downslope, the gradients decreasing from oldest to youngest. The oldest glacis are often only present as narrow residual ridges or outlying mesas, as planation of lower glacis have progressively removed upper glacis. Coque and Jauzein (1967) suggest that the number of glacis in Tunisia decreases systematically towards the south (Plate 53). Five glacis are present around Tunis and the High Steppe, south of Gafsa there are only four, the highest being present only as a few outliers.

Glacis d'érosion are thought to be erosional surfaces formed by fluvial action, cutting sequences of rocks that are easily eroded relative to the rocks of the adjacent upland. Supporting evidence for this fluvial model comes from the fact that stepped glacis are often paired on either side of contemporary channels rather like paired river terraces, and the fact that glacis are almost always covered with a layer of alluvium. This alluvial cover can be up to 15 m thick, though it rarely exceeds 10 m. Lower (younger) glacis tend to have thinner alluvial cover, and the alluvium tends to decrease in thickness towards the distal edge of the piedmont. The alluvium tends to be poorly sorted at the top of the glacis, becoming

Plate 53 A series of stepped glacis d'érosion developed on the southern flank of Djebel Sehib, southern Tunisia

better sorted downslope. In more arid areas, the alluvial cover of the glacis is frequently cemented by calcium carbonate or gypsum, forming an indurated DURICRUST. The role of duricrust in the formation of glacis is uncertain, although it may play an important part in preservation of older glacis.

Coque (1962) ascribes the formation of glacis in North Africa to slope retreat resulting from climate change; specifically a succession of humid and arid phases known to have affected the Sahara Desert during the Quaternary. He envisages a sequence of lateral planation during a humid phase, when moisture was sufficient to produce enough debris to balance the carrying capacity of streams, allowing them to erode laterally. This was followed by incision during an arid phase, when downcutting was promoted by lower sediment load in the streams. A return to more humid conditions resulted in renewed lateral planation at a lower level, forming a new glacis inset within the one above. This model is a gross oversimplification of the COMPLEX RESPONSE which river channels are now known to exhibit in response to environmental changes, but it is still generally believed that changes in the fluvial system resulting from climate changes are the basic trigger for formation of stepped glacis. The fact that the stepped glacis converge downslope indicates that changes in base level are unlikely to be involved in their formation. The widespread distribution of glacis across areas of different structural setting also rules out NEOTECTONICS as a major factor in their formation.

References

Coque, R. (1960) L'evolution des versants en Tunisie présaharienne, *Zeitschrift für Geomorphologie Supplementband* 1, 172–177.

Coque, R. (1962) *La Tunisie Présaharienne. Etude Géomorphologique*, Paris: Armand Colin.

Coque, R. and Jauzein, A. (1967) The geomorphology and Quaternary geology of Tunisia, in L. Martin (ed.) *Guidebook to the Geology and History of Tunisia*, 227–257, Tripoli: Petroleum Exploration Society of Libya.

Dresch, J. (1957) Pediments et glacis d'érosion, pédiplains et inselbergs, *Information Géographique* 22, 183–196.

Oberlander, T.M. (1989) Slope and pediment systems, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 56–84, London: Belhaven.

Plakht, J., Patyk-Kara, N. and Gorelikova, N. (2000) Terrace pediments in Makhtesh Ramon, central Negev, Israel, *Earth Surface Processes and Landforms* 25, 29–39.

White, K. (1991) Geomorphological analysis of piedmont landforms in the Tunisian Southern Atlas using ground data and satellite imagery, *Geographical Journal* 157, 279–294.

SEE ALSO: alluvial fan; desert geomorphology; pediment

KEVIN WHITE

GLACITECTONIC CAVITY

Glacitectonic cavities are narrow and sub-horizontal openings generated in bedrock by traction under a flowing GLACIER (Schroeder *et al.* 1986). The parallel walls take the form of irregular chevrons that follow the vertical joint pattern, while the roofs follow stratification planes. In some cases, well-compacted lodgement till forms the roof. The irregular floor of galleries is usually covered by debris issued from localized roof or walls failures.

Located less than 20 m below the surface, glacitectonic cavities can be hundreds of metres long, but typically less than 3 m wide and less than 10 m high. As their presence is only revealed by chance, from excavation work or local roof failures, they constitute hazardous constraints in URBAN GEOMORPHOLOGY, especially in eastern Canada (Schroeder 1991).

Glacitectonic cavities are found below planar topographies, within sub-horizontal limestone or thinly bedded shale. Movement and weight of a flowing inlandsis, possibly aided by dissolution along the stratification planes, allows rock sheets to slide one on the other, leading to the spreading apart of vertical joints and to the creation of glacitectonic voids.

References

Schroeder, J. (1991) Les cavernes à Montréal, du glaciotectonisme à l'aménagement urbain, *Canadian Geographer* 35(1) 9–23.

Schroeder, J., Beaupré, M. and Cloutier, M. (1986) Ice-push caves in platform limestones of the Montreal area, *Canadian Journal of Earth Sciences* 23, 1,842–1,851.

JACQUES SCHROEDER

GLACITECTONICS (GLACIOTECTONICS)

Glacitectonic deformation may be defined as 'the structural deformation as a direct result of glacier movement or loading' (INQUA Work Group on Glacier Tectonics 1988). This term was first introduced by Slater (1926), and re-examined by Banham (1975). This topic has been studied a great deal in recent years, and a number of collections of papers on the subject have been published (Aber 1993; Warren and Croot 1994) and an online bibliography (http://www.emporia.edu/s/www/earthsci/biblio/biblio.htm). Additionally, recent textbooks on glacial geology include detailed sections on glacitectonic deformation (Benn and Evans 1997; van der Wateren 1995).

However, prior to the 1980s, glacitectonic deformation was thought to be a rare phenomenon, and was studied as a distinct field in glacial sedimentology. This view was first challenged when Boulton and Jones (1979) suggested that a significant proportion of glacier motion occurred not in the ice, but in a saturated weak deforming layer beneath the ice. These results showed how glacitectonic deformation was an integral part of the glacial environment, and not an unusual occurrence.

There are two types of glacitectonic deformation formed by the action of a moving glacier (Figure 76):

(a) *Proglacial deformation* which takes place at the glacier margin and is characterized by pure shear and compressional tectonics, i.e. open folds, thrusts and nappes. This results in the formation of push moraines;
(b) *Subglacial deformation* which takes place beneath the glacier and is characterized by simple shear and extensional tectonics, i.e. attenuated folds, boudins and augens, and results in the formation of deformation till and/or flutes and drumlins.

Similar styles of deformation can also occur within the ice itself (Hart 1998) as well as within permafrost (Astakhov *et al.* 1996).

Proglacial glacial tectonic structures have been relatively well studied because of their accessibility. In fact, the large number of studies of these features led many workers in the past to consider only proglacial deformation in discussions of glacitectonics. In contrast, subglacial deformation has had the least study because of the logistical

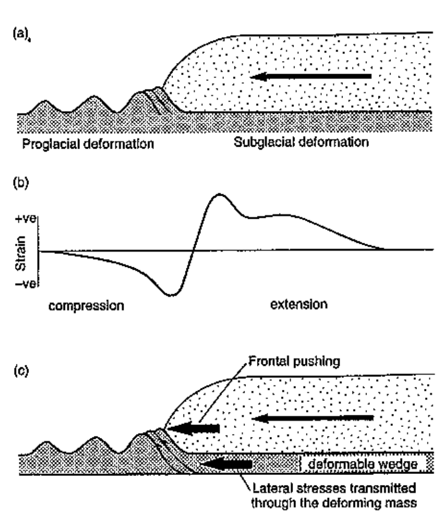


Figure 76 (a) schematic diagram showing the positions of subglacial and proglacial glacitectonic deformation; (b) the theoretical pattern of longitudinal strain; (c) schematic diagram of the forces producing proglacial deformation: frontal pushing and compressive stresses transmitted through a subglacial deformable wedge (after Hart 1998) problems involved in subglacial process studies, but the number of studies has dramatically risen in the past ten years.

Additionally, deformation also occurs within the glacial environment as a result of gravitation instabilities associated with stagnant ice and is known as dead-ice tectonics. Typical features include ice-collapse structures in an outwash plain, debris-flow mobilizations of till, and instability of subglacial sediments to produce 'crevasse infill' structures. These features are not glacitectonic structures *sensu stricto*, but may reflect the presence of saturated till in the subglacial environment, and so may be associated with subglacial deformation.

Proglacial glacitectonic deformation

Proglacial glacitectonic deformation is generally characterized by large-scale compressional folds and thrusts. The usual result of the proglacial deformation processes is to produce a topographic ridge transverse to the ice margin called a push moraine. There is often a basin up-glacier from which the material of the ridge has been removed. However, they do not always have a topographic expression. Many proglacial structures have been subsequently overridden by ice and so have become incorporated into drift deposits and have little or no topographic expression.

Push moraines are very common and can occur on scales ranging in height from 0.5 to 50 m, and in length from 1 m to several kilometres. Push moraines are associated with both contemporary glaciers and Quaternary glaciations (as well as pre-Quaternary glaciations) (see reference list). A recent review of push moraines is by Bennett (2001).

It has been argued by numerous workers that proglacial deformation can be modelled as thin-skin thrust tectonics, and the processes involved in the formation of push moraines are similar to mountain building in hard rock tectonic terrains. Using the work of Hubbert and Rubey (1959), many researchers have argued that glacitectonic nappes move along incompetent rock units or planes of weakness due to high pore-water pressures.

Although there are many processes associated with proglacial glacitectonic deformation, they can be generally divided into two types:

1. *'Foreland only' deformation* Where there is no deforming bed present, deformation may only take place in the foreland by the deformation of pre-existing sediments. This may typically include sandur sediments in terrestrial environments and shallow marine or fjord sediments in marine environments.
2. *'Deformable wedge' deformation* Where there is a deforming bed present, the subglacial and proglacial environment can be modelled as a deformable wedge which deforms by gravity spreading driven by the ice (Figure 76c). Deformation occurs due to both down-ice thrusting of the glacier into the foreland, as well as the horizontal component of the glacier's effective pressure (normal pressure minus pore-water pressure) transmitted through the subglacial layer into the foreland.

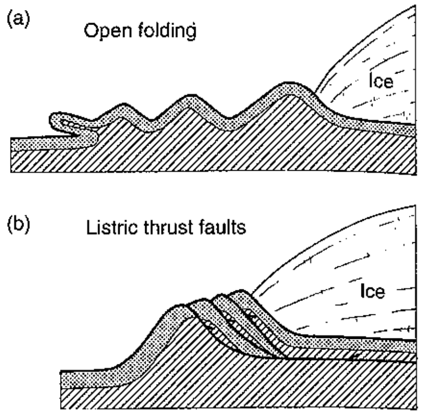


Figure 77 Schematic diagrams of different proglacial push moraines: (a) open folding; (b) listric thrust folding (after Hart and Boulton 1991)

Deformation of sediments (in both styles of push moraine) range from ductile (open folding) to brittle (thrust faulting and thrust nappes) (Figure 77). These styles of deformation reflect both the competancy of the material and increasing longitudinal compression from the simple folding to more complex nappe structures. Deformation structures are also found at the base of thrust faults and nappes, with tectonic breccias associated with brittle deformation and shear zones formed associated with ductile deformation.

Subglacial glacitectonic deformation

Although there are fewer studies of the subglacial environment due to its inaccessibility, there is still a considerable body of literature on subglacial deformation. Early descriptions of deformation structures in till included folds, laminations and blocks, boudins or rafts of soft sediments; such till was called 'deformation till'.

Subglacial deformation can occur beneath warm-based glaciers, when meltwater released from the glacier bed cannot easily escape from the system, so that pore-water pressures in the subglacial sediments build up and sediment strength is reduced:

$$\tau = (P_i - P_w) \tan \varphi$$

where τ is basal friction, P_i is ice overburden pressure, P_w is pore-water pressure and $\tan \varphi$ is the angle of internal friction (Coulomb's Law).

STUDY METHODS

In recent years subglacial deformation has been studied by three ways: (1) *in situ* process studies; (2) geophysical techniques; and (3) sedimentology. These methods have been discussed in detail in Hart and Rose (2001).

In situ subglacial process studies consist of the monitoring of the subglacial environment by inserting instruments into the subglacial bed via hot water drilled boreholes. This is a relatively simple technique and has been used on about ten modern glaciers including Breiðamerkurjökull (Iceland), Ice Stream B (Antarctica), Trapridge Glacier (Canada), Black Rapids Glacier (Alaska), Storglaciären (Sweden) and Bakaninbreen (Svalbard). These studies reveal the average thickness of the deforming layer is 0.5 m and indicate that deformation does occur beneath most of these glaciers. However, the importance of the effects on basal motion due to subglacial deformation ranges between 100 per cent at Black Rapids Glacier, Alaska to 13 per cent at Ice Stream B, Antarctica. Although the reason for the difference is not yet known, it has been suggested that the granulometry of the subglacial sediment may account for the difference in behaviour within the deforming layer. The glaciers with coarse-grained till appear more likely to have a higher percentage of basal motion due to sediment deformation, whilst those with more clay-rich lithologies may have only very thin deforming layers.

In addition, the presence of a deforming bed over large areas has been identified by seismic investigations in Antarctica, in particular beneath Ice Stream B and the Rutford Ice Stream.

However, most studies of subglacial deformation have been based on sedimentological studies from both modern and Quaternary glacial sequences. Most researchers have argued that the subglacial deforming layer behaves as a shear zone, which is a narrow band of high overall ductile shear located between sub-parallel walls. This deformation results in three features that will be discussed briefly below: deformation till, deformation structures and subglacial bedforms.

Deformation till

Hart and Boulton (1991) have argued that the resultant till from subglacial deformation is deformation (or deforming bed) till, which is a primary till formed from a combination of both deformation and deposition. It forms by direct melt-out of debris from the ice above, advection from till up-glacier, and changes in the thickness of the deforming layer. Where layers of deformation till are accreted on one another this is known as constructional deformation. In contrast, where the deforming layer thickens (due to changes in effective pressure, or large rafts of bedrock being thrusted into the shear zone), this is known as excavational deformation.

Deformation structures

Features typical of shear zones include: folds, boudins, augens, rotated clasts and tectonic laminations (Plate 54). The latter form from the attenuation of perturbations of the base of the deforming layer, producing ungraded laminations. However, these features will only be visible if they are formed from the mixing of sediments with different lithologies or competencies, under relatively low to medium simple shear. At very high shear strains these tectonic features can become homogenized and so macro-scale structures may not be visible. Instead criteria such as a specific till fabric (low strength associated with a thick deforming layer, high strength associated with a constrained deforming layer), or specific micromorphology structures (evidence of rotation or shears) may be used as a distinguishing criterion.

Subglacial bedforms

It has also been argued that subglacial streamlined bedforms (lineations, flutes and drumlins) are a product of subglacial deformation (Boulton

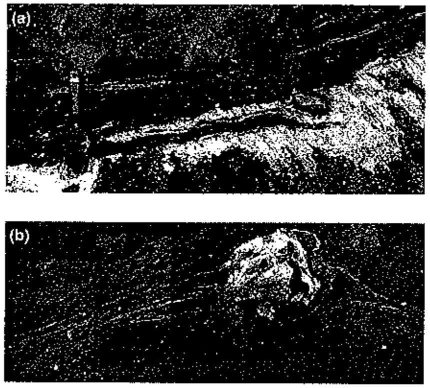


Plate 54 Examples of subglacial deformation: (a) chalk being attenuated to form tectonic laminations, West Runton, Norfolk; (b) chalk laminations flowing around on obstacle (flint clast), Weybourne, Norfolk (photographs by Kirk Martinez)

1987). These features form due to the presence of more competent masses (or cores) within the deforming layer which act as obstacles to flow. Where the core of the drumlin is weak then deformation structures will be seen, but these are relatively rare, and instead most drumlins have a competent core and a carapace composed of deformation till.

Using this sedimentological data, a number of authors have argued for wide spread subglacial deformation beneath the Pleistocene glaciers where the ice moved over the unconsolidated rocks of the European (and British) and Laurentide ice sheets.

Conclusions

Glacitectonic deformation is a fundamental process in glacier behaviour and a key component in the proglacial and subglacial sediment/ice deposition, erosion and transport system. There are very few modern glaciers that do not show evidence for proglacial deformation at their margins, and subglacial *in situ* process studies have revealed that subglacial deformation is also a common process. In addition, studies of Quaternary sediments demonstrate that such processes were also widespread in the past.

As a result, any study of the glacial environment needs to take glacitectonics into consideration, and future research needs to focus on the geotechnical properties of till to further understand the links between sediment behaviour and ice dynamics.

References

Aber, J.S. (ed.) (1993) *Glaciotectonics and Mapping Glacial Deposits*, Canadian Plains Research Centre, University of Regina.

Astakhov, V.I., Kaplyanskaya, F.A. and Tarnogradsky, V.D. (1996) Pleistocene permafrost of West Siberia as a deformable glacier bed, *Permafrost and Periglacial Processes* 7, 165–191.

Banham, P.H. (1975) Glaciotectonic structures: a general discusion with particular reference to the Contorted drift of Norfolk, in A.E. Wright and F. Moseley (eds) *Ice Ages, Ancient and Modern*, 69–94, Liverpool: Seel House Press.

Benn, D.I. and Evans, D.J.A. (1997) *Glaciers and Glaciation*, London: Arnold.

Bennett, M.R. (2001) The morphology, structural evolution and significance of push moraines, *Earth-Science Reviews* 53, 197–236.

Boulton, G.S. (1987) A theory of drumlin formation by subglacial deformation, in J. Menzies and J. Rose (eds) *Drumlin Symposium*, 25–80, Rotterdam: Balkema.

Boulton, G.S. and Jones, A.S. (1979) Stability of temperate ice caps and ice sheets resting on beds of deformable sediment, *Journal of Glaciology* 24, 29–44.

Hart, J.K. (1998) The deforming bed/debris-rich basal ice continuum and its implications for the formation of glacial landforms (flutes) and sediments (melt-out till), *Quaternary Science Reviews* 17, 737–754.

Hart, J.K. and Boulton, G.S. (1991) The interrelationship between glaciotectonic deformation and glaciodeposition, *Quaternary Science Reviews* 10, 335–350.

Hart, J.K. and Rose, J. (2001) Approaches to the study of glacier bed deformation, *Quaternary International* 86, 45–58.

Hubbert, M.K. and Rubey, W.W. (1959) Role of fluid pressure in mechanics of overthrust faulting, *Geological Society of America Bulletin* 70, 115–166.

Slater, G. (1926) Glacial tectonics as reflected in disturbed drift deposits, *Geologists' Association Proceedings* 37, 392–400.

Warren, W.P. and Croot, D.G. (eds) (1994) *Formation and Deformation of Glacial Deposits*, Rotterdam: Balkema.

Wateren, van der F.M. (1995) Processes of glaciotectonism, in J. Menzies (ed.) *Modern Glacial Environments: Processes, Dynamics and Sediments*, 309–333, Oxford: Butterworth-Heinemann.

Further reading

Croot, D.G. (ed.) (1988) *Glaciotectonics: Forms and Processes*, Rotterdam: Balkema.

JANE K. HART

GLINT

A marked gemorphological line dividing (the Canadian and the Baltic) SHIELDs from neighbouring stable platforms (the Great Plains and the Russian Plains, respectively) in the northern hemisphere. It is manifest in an ESCARPMENT which extends hundreds of kilometres and rises 20–100 m above the shield. The front of this escarpment is called a glint line. Pre-Pleistocene DENUDATION and, more significantly, differential scouring of expanding ICE SHEETs during the Pleistocene is responsible for glint formation. The Palaeozoic (Ordovician, Silurian) limestones, dolomites and sandstones of the tableland are more resistant to glacial erosion than the weathered Precambrian igneous and metamorphic rocks of the shield. Forward pushing ice was temporarily halted by the escarpment and thus allowed deep scouring at its base. After ice retreat meltwater accumulated in the depressions and glint lakes originated.'

Glint is an Estonian term of Germanic origin. Once it denoted cliffs along coasts. The Baltic–Ladoga Glint extends from the islands of Estonia along the southern coast of the Gulf of Finland. Lake Ladoga and Lake Oniega occupy the depressions. Although the term is not in use in Canada and the United States, the glint line is also present there. Some major lakes, including the Great Bear, the Great Slave, Lake Winnipeg and the Great Lakes are all glint lakes, a subclass of ice-scoured lakes. Niagara Falls is the best known example of waterfalls along the glint line.

DÉNES LÓCZY

GLOBAL GEOMORPHOLOGY

The term global geomorphology embodies the notions of studying landform development at large spatial and temporal scales, of emphasizing global variations in landforms and geomorphic processes, of investigating the interactions between the land surface and other components of the Earth system, and of appreciating the particular combination of conditions for landform genesis on Earth compared with the other solid planetary bodies of the Solar System.

In focusing attention on large-scale phenomena and change over long periods of time, global geomorphology is concerned primarily either with the development of very large individual landform features, such as an entire mountain range, or with the assemblage of smaller individual landforms making up whole landscapes. At these large spatial and temporal scales the internal geomorphic processes of volcanism and tectonics generally become more significant in relation to surface geomorphic processes. A further consequence of looking at the macroscale is that short-term measurements of geomorphic processes on their own provide relatively little insight into the nature and rates of processes responsible for landscape genesis. Although numerical models (see MODELS) have been developed to investigate large-scale landscape change, data relevant to testing such models generally have to pertain to periods of thousands to millions of years, rather than the timescale of years or decades of modern process measurements.

A methodological consequence of these long timescales is that the approach to global geomorphology is predominantly historical where the emphasis is on explaining the conditions and processes responsible for development over time of a single major landform, or a regional or larger scale landscape. This contrasts with the dominance of the functional approach in small-scale surface process geomorphology where the main interest is understanding the adjustment of form to process over short periods of time.

Another distinction between global geomorphology and smaller scale approaches to landform analysis is the frame of reference that must often be employed. At the small scale it is usually sufficient to know local slope gradients and height differences, such as between interfluves and river channels, rather than absolute changes in elevation with respect to sea level. In global geomorphology, however, constraining changes in absolute elevation of the land surface above sea level is required in order to relate changes in regional topography through time to rates of crustal uplift and rates of denudation.

Historical context

A global approach to geomorphology is not new. An important theme in the study of the landforms up to the nineteenth century was the attempt to understand the origin and history of the surface of the Earth as a whole. For instance, one of the elements of Charles Lyell's concept of UNIFORMITARIANISM, was the notion of a steady-state Earth in which uplift in some areas was 'balanced' by subsidence in others, and where changes in the location of uplift and subsidence occurred over time, but the overall form of the Earth's surface did not change substantially.

Lyell's idea of regions of crustal uplift and subsidence was taken up by Charles Darwin who, during the earlier part of his career dominated by writings on geological subjects, sought to develop a global synthesis relating uplift of the continents to processes such as volcanism and mountain building. More specifically, Darwin adopted Lyell's notion of regions of subsidence in developing his own theory of coral atoll formation in which he envisaged coral reefs growing upwards from the substrate of volcanic islands grouped in broad regions of what he inferred was subsiding ocean crust. In developing his coral reef theory, Darwin provided an object lesson in historical methodology applied to understanding landform development by suggesting how spatial patterns of a range of related landforms – in this case, volcanic islands, barrier reefs, fringing reefs and atolls – could, with careful observation and reasoning, be viewed as representing stages in the development of a single landform through time.

This strategy was taken up and extended by William Morris Davis who, in his CYCLE OF EROSION, sought to develop a general evolutionary scheme for landscape development, where the form of a landscape was seen as a product of the rock structures present and the surface geomorphic processes operating, but predominantly as a function of stage of development. Although acknowledging complications arising from variations in climatic conditions, and particularly from intermittent crustal uplift, Davis's evolutionary model depended heavily on the reality of distinctive landscapes being created at different stages of the cycle of erosion. In its simplest form this involved rapid uplift from close to sea level of a low relief land surface, its progressive incision by river systems creating maximum local relief, and then an extended period of interfluve lowering with respect to valley bottoms until a low relief surface close to sea level, or PENEPLAIN, was restored.

Although often heavily dependent on particular interpretations of landscape features, and thus far less secure than Darwin's earlier exemplary treatment of coral atoll formation, the evolutionary approach advocated by Davis became the dominant strategy of global geomorphology through the first half of the twentieth century, at least amongst geomorphologists in Britain and North America. In Germany a different model of landscape change was developed by Walther Penck who emphasized the importance of the interplay between external erosional processes and internal tectonic processes causing uplift. Although more in sympathy with modern approaches to understanding large-scale landscape development, Penck's more complex approach to landform analysis never achieved the influence of the simple version of Davis's evolutionary model.

The idea of the development over millions of years of extensive, low relief erosion surfaces graded to sea level led to the development of DENUDATION CHRONOLOGY, a method of landscape analysis in which low relief erosion surfaces at different elevations were interpreted in terms of falls in base level resulting either from eustatic sea-level change (global sea-level fall), or from tectonic uplift of the land surface. Throughout the mid-twentieth century much emphasis was placed in correlating supposed remnants of particular surfaces considered to have resulted from specific uplift events. Taken to its ultimate extent, the correlation of such erosion surface remnants across the continents was seen as potentially being a chronological replacement for stratigraphy where the sedimentary record was absent or incomplete. The fullest development of denudation chronology as a methodological basis for global geomorphology is perhaps to be found in the work of Lester King who interpreted flights of low relief surfaces across different continents as representing synchronous episodes of continental uplift of global extent.

The decline in interest in global geomorphology and the corresponding move towards studies of small-scale surface process geomorphology from the 1950s occurred for two main reasons. One was the wholly inadequate dating control that was usually available for the denudation chronologies presented, the other a lack of understanding of the tectonic and surface geomorphic processes that could create erosion surfaces graded to sea level, and then subsequently preserve remnants of them when base level fell.

Renewed interest in global geomorphology

Although the revolution in Earth sciences arising from PLATE TECTONICS might have been expected to reinstate interest in global geomorphology, little attention was paid by most geomorphologists to this integrative global-scale model when it was formulated in the late 1960s and early 1970s, presumably because the focus by that time was on quantitative approaches to small-scale surface geomorphic processes. A renewed concern with global geomorphology is really only evident from the 1980s, and it occurred for a number of reasons. One was the growing availability of satellite remote sensing imagery which made evident the large-scale components of the Earth's landforms. Although initially used primarily to explore regional and subcontinental-scale landform associations, by the 1990s satellite data was being used to create digital elevation models (DEM) of the land surface at horizontal resolutions down to a few metres. This added to the growing number of digital topographic data sets being created from national archives of topographic data. By 2001 the Shuttle Radar Topography Mission had collected high resolution radar-based elevation data covering the Earth's surface between latitudes ~60°S to 60°N.

At the same time that Earth-orbiting satellites were providing images of terrestrial landscapes, there was a flood of remote sensing imagery from planetary missions, such as the Viking missions to Mars and the Voyager missions to the outer planets in the 1970s, the Magellan mission to Venus in the 1980s and the Mars Global Surveyor which provided high resolution images. Understanding of the tectonics, volcanism, surface processes and climatic history of these planetary bodies relied heavily on the interpretation of their landforms, where possible by comparisons with supposed terrestrial analogues (for instance, the outflow channels on Mars which were seen to have many similarities to the landscapes of catastrophic flooding in the Channeled Scabland of eastern Washington, USA). At the same time, the scale of landforms seen in planetary imagery pointed to the insights to be gained by studying terrestrial forms with a similar global perspective (see EXTRATERRESTRIAL GEOMORPHOLOGY).

Another important reason for increased interest in global geomorphology was the development of the computing capability necessary to numerically model regional-scale landscapes over geological time spans. Although small catchment/channel slope-scale surface process models have been developed by geomorphologists and hydrologists since the 1960s, numerical models of regional-scale landscape evolution incorporating tectonic deformation and isostasy, as well as surface processes, have been under active development only since the late 1980s.

Constraining such models requires data on denudation rates for time spans of millions of years relevant to long-term landscape development. The increasing availability of such information from the 1980s as a result of new geochronological methods and data sources is another reason for the revived interest in global geomorphology. Hydrocarbon exploration along continental margins has provided a wealth of data on rates of sediment deposition from which denudation rates on the adjacent hinterland can be estimated, at least where the sediment source area and its changes over time can be constrained. More important, however, has been the application of thermochronological techniques to infer denudational histories and denudation rates. A range of low-temperature techniques such as $^{39}Ar/^{40}Ar$ dating, fission-track thermochronology (see FISSION TRACK ANALYSIS) and helium thermochronology can now provide information on the cooling history of rocks in the upper few kilometres of the Earth's crust. This information on the timing and rate of cooling can be converted into estimates of denudation rates averaged over periods of millions of years since it is the progressive stripping of crust by denudation that is largely responsible for shallow crustal cooling. These data provide information on broad regional patterns of denudation, but they can now be related to more local denudation rates by coupling with data from cosmogenic isotope analysis (see COSMOGENIC DATING) which provides denudation rates over timescales of thousands to hundreds of thousands of years.

Key issues in global geomorphology

The most obvious issue in global geomorphology is to understand the gross variations in the Earth's continental topography and how this topography has changed over time. Why, for instance, is 82 per cent of the world's land surface over 4,000 m above sea level concentrated in the Tibetan Plateau? And what is the origin of the large area of anomalously high topography extending across southern Africa and into the adjacent Atlantic Ocean. Answers to these questions require an understanding of the interaction of internal and external processes over periods of millions of years. Crucial to answering such questions are data on changes on the elevation of the land surface over time, since the timing of the uplift of the Tibetan Plateau, for instance, is key to understanding the cause of such uplift.

Unfortunately, constraining such surface uplift has proved very difficult, not least because denudation in uplifted terrain tends to remove evidence that would be indicative of prior elevations. However, various techniques have been developed to infer past elevations in addition to the obvious strategy of using shoreline or shallow marine deposits where present. These include inferring temperature (and therefore indirectly elevation) change from the characteristics of fossil leaves on the basis that specific plant types have particular temperature tolerances and that surface uplift will elevate fossils into cooler climatic zones. This approach has been used to infer surface uplift in mountain ranges such as the Himalayas, but it requires detailed information on global and regional climatic changes which would also produce vertical shifts in climatic zones. Another approach is to use the elevation-dependent fractionation of oxygen in precipitation across mountain ranges which can be incorporated into carbonate sediments, but of considerable potential is basalt vesicle ratio analysis. This technique uses the effect of atmospheric pressure of the relative size of gas bubbles at the top and bottom of individual lava flows to infer the atmospheric pressure, and hence elevation, at the time of eruption. Notwithstanding these and other techniques, constraining changes over time in the absolute elevation of the land surface remains a problematic but fundamental issue in global geomorphology.

Another important issue is the coupling of onshore and offshore records of denudation and deposition. The growth in offshore hydrocarbon exploration along continental margins since the 1970s has greatly expanded our knowledge of their depositional history, but it has also raised the question of what controls the supply of sediment from the adjacent continental hinterland. Answering this question requires information not just on the mobilization and transport of sediment from onshore to offshore but also on tectonic mechanisms and the isostatic response to changes in crustal loading as mass is transferred offshore. Although largely irrelevant to small-scale surface process geomorphology, ISOSTASY assumes a critical role in global geomorphology since, at these larger spatial and temporal scales, flexure of the lithosphere in response to denudational unloading can have important effects on the mode of landscape development.

A further key theme in global geomorphology is the coupling between internal and external processes. Although the influence of tectonic mechanisms on surface processes through the construction of relief has been long understood, the way in which spatial variations in denudation rates can affect patterns of tectonic deformation was only fully appreciated in the 1990s. This is evident in the commonly found strike-parallel pairing of metamorphic facies in mountain ranges as a result of higher rates of denudation (and therefore greater depths of exposure) on the wetter, windward side compared with the drier, leeward side. Modelling of patterns of crustal deformation as a result of spatial variations in denudation rates has further emphasized the two-way interaction between surface and internal geomorphic processes.

The role of the land surface in interactions between tectonics and climate has also received attention in attempts to understand the long-term geological controls over the concentration of atmospheric carbon dioxide and hence, through the greenhouse effect, global climate. The key process here is the weathering of silicate minerals, a reaction which draws down CO^2 from the atmosphere. As global topography and relief changes as a result of interactions between tectonics, climate and landscape development, the global rate of CO^2 draw-down would be expected to vary, although the operation of these interactions are far from fully understood.

Finally, comparative planetary geomorphology provides the key perspective for global geomorphology. Looking at landscape development on other planetary bodies shifts our perspective from viewing terrestrial landforms as 'normal', and emphasizes that the Earth's landforms have arisen from a particular combination of its size, its composition, its distance from the sun, the composition and density of its atmosphere, and its age. The great majority of planetary bodies have surfaces dominated by impact craters, making impact cratering the dominant geomorphic process in the Solar System (although most occurred in the first 500 to 600 Ma from the birth of the Solar System around 4.5 Ga ago). The critical factor for Earth is the surface temperatures that it experiences which encompass the range over which water can exist as a solid, a liquid and a gas. This enables Earth to have an active hydrological cycle which is key to many geomorphological processes. Also Earth's size and composition means that it has a high enough internal temperature to melt rock and therefore permit volcanism and the convection that helps power plate tectonics. Earlier in its history, Mars also probably experienced short-lived episodes when there was a fairly active hydrological cycle including oceans; this is the main period of the channel formation on Mars. By contrast, the high surface temperatures on Venus, largely resulting from an intense greenhouse effect associated with a dense CO^2-rich atmosphere, has prevented the existence of liquid water; thus its surface is dominated by the effects of volcanism and impact cratering.

Further reading

Burbank, D.W. and Anderson, R.S. (2001) *Tectonic Geomorphology*, Malden, MA: Blackwell Science.

Ellis, M. and Merritts, D. (1994) *Tectonics and Topography*, Washington, DC: American Geophysical Union.

Greeley, R. (1994) *Planetary Landscapes*, 2nd edition, London: Chapman and Hall.

Stüwe, K. (2002) *Geodynamics of the Lithosphere*, Berlin: Springer-Verlag.

Summerfield, M.A. (1991) *Global Geomorphology*, London: Longman.

——(ed.) (2000) *Geomorphology and Global Tectonics*, Chichester: Wiley.

MIKE SUMMERFIELD

GLOBAL WARMING

There is now a widespread appreciation that the build-up of greenhouse gases in the atmosphere (carbon dioxide, methane, nitrous oxide, CFCs, etc.) will create an enhanced greenhouse effect that will cause global warming. Details of the degree of warming that will occur and of the associated changes in other climatic variables are provided in the reports of the Intergovernmental Panel on Climate Change (2001). If such changes occur over coming decades certain landscapes and geomorphological processes will be modified (Table 22).

Some landscapes, 'geomorphological hot spots', will be especially sensitive because they are located in zones where it is forecast that climate will change to an above average degree. In the high latitudes of Canada or Russia the degree of warming may be three or four times greater than the global average. It may also be the case with respect to some critical areas where particularly substantial changes in precipitation may result from global warming. For example, various scenarios suggest the High Plains of the United States of America will become markedly drier. Other landscapes will be highly sensitive because certain landscape-forming processes are so closely controlled by climatic conditions. If such landscapes are close to particular climatic thresholds then quite modest amounts of climatic change can flip them from one state to another. In this entry attention will be paid to some of these hot spots.

Tundra and permafrost terrains

High latitude tundra and PERMAFROST terrains may be regarded as one of these sensitive zones. They are likely to undergo especially substantial temperature change. In addition, the condition of permafrost is particularly closely controlled by temperature conditions. By definition it cannot occur where mean annual temperatures are positive, and the latitudinal limits of different types of permafrost can be related to varying degrees of negative temperatures. Thus the equatorward limit of continuous permafrost may approximate to the −5 °C isotherm and the equatorward limit of discontinuous or sporadic permafrost to the −2 °C isotherm. It is likely that the latitudinal limits of permafrost will be displaced polewards by 100 to 250 km for every 1 °C rise in mean annual temperature. The quickest loss of permafrost would occur in terrains underlain by surface material with low ice contents. The slowest response would be in ice-rich materials, which require more heat to thaw. Snow or the presence of thick, insulating organic layers (i.e. peat) might also buffer the effects of increased surface temperatures in some areas.

There is historical evidence that permafrost can degrade speedily. For instance, during the warm 'optimum' of the Holocene (*c*.6,000 years ago) the southern limit of discontinuous permafrost in the Russian Arctic was up to 600 km north of its present position (Koster 1994). Similarly, researchers have demonstrated that along the Mackenzie Highway (Canada), between 1962 and 1988, the southern fringe of the discontinuous zone had moved north by about 120 km in response to an increase over the same period of 1 °C mean annual temperature (Kwong and Tau 1994).

Table 22 Some geomorphologic consequences of global warming

Hydrologic
Increased evapotranspiration loss
Increased percentage of precipitation as rainfall at expense of winter snowfall
Increased precipitation as snowfall in very high latitudes
Possible increased risk of cyclones (greater spread, frequency and intensity)
Changes in state of peatbogs and wetlands
Less vegetational use of water because of increased CO_2 effect on stomatal closure
Vegetational controls
Major changes in latitudinal extent of biomes
Reduction in boreal forest, increase in grassland, etc.
Major changes in altitudinal distribution of vegetation types (c.500 m for 3 °C)
Growth enhancement by CO_2 fertilization
Cryospheric
Permafrost, decay, thermokarst, increased thickness of active layer, instability of slopes, river banks, and shorelines
Changes in glacier and ice-sheet rates of ablation and accumulation
Sea-ice melting
Coastal
Inundation of low-lying areas (including wetlands, deltas, reefs, lagoons, etc.)
Accelerated coast recession (particularly of sandy beaches)
Changes in rate of reef growth
Spread of mangrove swamp
Aeolian
Increased dust storm activity and dune movement in areas of moisture deficit
Soil erosion
Changes in response to changes in land use, fires, natural vegetation cover, rainfall erosivity, etc.
Changes resulting from soil erodibility modification (e.g. sodium and organic contents)
Subsidence
Desiccation of clays under conditions of summer drought

Woo *et al.* (1992) made certain predictions based on the assumption that a greenhouse warming of 4–5 °C causes a spatially uniform increase in surface temperature of the same magnitude over northern Canada. They suggested that permafrost in over half of what is now the discontinuous zone could be eliminated, that the boundary between continuous and discontinuous permafrost might shift northwards by hundreds of kilometres and that a warmer climate could ultimately eliminate continuous permafrost from the whole of the mainland of North America, restricting its presence only to the Arctic Archipelago.

In areas where rapid permafrost melting occurs, the consequences will be legion. They include ground subsidence (THERMOKARST), increased erosion of shorelines and riverbanks, and an increase in debris flow activity and other forms of slope instability.

High latitude areas may also be particularly susceptible to changes in precipitation and runoff. Areas which are currently very dry, because the air is so cold, may become moister as warmer winters cause more snow to fall, thereby creating a likelihood of increased summer runoff. In somewhat warmer environments, where substantial winter snowfall occurs, there might be a tendency in a warmer world for a decrease in the proportion of winter precipitation that falls as snow. There would thus be greater winter rainfall and runoff, but less overall precipitation to enter snowpacks to be held over until spring snowmelt. This in turn would have adverse consequences both for late spring and summer runoff levels in rivers and for soil moisture levels. Other factors may also modify runoff. For example, as permafrost thaws, groundwater recharge may increase and surface runoff decrease.

Glaciers and ice sheets

Glaciers and ice sheets will be highly susceptible to a rise in temperature. Although there has been considerable debate as to whether or not polar ice caps might respond catastrophically to global warming because of an increase in ablation, accelerated melting of tidewater snouts, the cliffing of termini by a rising sea level, or the removal of the buttressing effects of ice shelves as they melt (Huybrechts *et al.* 1990), it is probably valley glaciers in alpine situations which will respond most quickly and markedly to climatic warming. Such glaciers are highly responsive, as is made evident by their frequent and rapid fluctuations during the Neoglacials of the Holocene. Although topographic controls and changes in precipitation and cloudiness are significant controls of glacier state, it is highly likely that most alpine glaciers will show increasing rates of retreat in a warmer world. Indeed, given the rates of retreat (20–70 cm year) experienced in many mountainous areas in response to the warming episode since the 1880s, it is probable that many glaciers will disappear altogether, from areas as diverse as the Highlands of East Africa or the Southern Alps of New Zealand.

Desert margins

The history of desert margins indicates that in the past they too have been sensitive to environmental change. This in turn suggests that they are likely to be susceptible to future environmental changes. Thus closed depressions have fluctuated repeatedly from being dry and saline to being full and fresh. Valley bottoms and hillsides have alternated between cut-and-fill, and dunefields have at some times been mobile and at other times stable (Forman *et al.* 2001). Many dry regions will suffer large diminutions in runoff (Arnell 1999), with annual totals likely to be reduced by over 60 per cent. Indeed Shiklomanov (1999) has suggested that in arid and semi-arid areas an increase in mean annual temperature by 1 to 2 °C and a 10 per cent decrease in precipitation could reduce annual river runoff by up to 40–70 per cent.

In the case of closed depressions, the dating of high water levels in lakes in the tropics and subtropics shows that many of them have had a complex history during the Holocene and that their water levels have varied considerably. High levels were a feature of the Saharan region around 8,000 years ago, a time when global temperatures were probably slightly greater than today. Very large numbers of freshwater deposits date from this time, even in the dry heart of the Sahara. Some stream courses (e.g. the Wadi Howar) were active.

In the case of river and slope systems, they too have fluctuated between phases of stability or alluviation and phases of erosion and incision. Even over the last century or so the valley systems of the American south-west, called ARROYOS, have experienced trenching and filling episodes in response to climatic and other stimuli (e.g. land use change). Of particular importance have been changes in the amount and intensity of precipitation. Crucial in this respect is the response of vegetation cover to rainfall events, for in semi-arid areas it is not only highly dependent on moisture availability but also controls the erodibility of the ground surface (Elliot *et al.* 1999).

Changes in precipitation and evapotranspiration rates also have a marked impact on aeolian environments and processes. Rates of deflation, sand and dust entrainment and dune formation are closely related to soil moisture conditions and vegetation cover. Areas that are at present marginal with respect to aeolian processes will be particularly susceptible, and this has been made evident, for example, through recent studies of the semi-arid portions of the United States (e.g. the High Plains). Repeatedly throughout the Holocene they have flipped from a state of vegetated stability to states of drought-induced surface instability. Thermoluminescent and optical dates have made evident their sensitivity to quite minor perturbations. Geomorphologists, using the output from General Circulation Models (GCMs), combined with a dune mobility index which incorporates wind strength and the ratio of mean annual precipitation to potential evapotranspiration, have shown that with global warming, sand dunes and sandsheets on the Great Plains are likely to become reactivated over a significant part of the region, particularly if the frequencies of wind speeds above the threshold velocity for sand movement were to increase by even a moderate amount (Muhs and Maat 1993). The same applies to dust storm generation in the Great Plains and the Canadian Prairies, where the application of GCMs shows that conditions comparable to the devastating dust-bowl years of the 1930s are likely to be experienced.

Tropical coastlines

Tropical coastlines are a further very sensitive environment with respect to future climatic change. This is for three main reasons: the relationship between tropical cyclone activity and the sea-surface temperature (SST), the temperature tolerances of coral reefs, and the effects that temperature change and sea-level rise have upon mangrove swamps.

Tropical cyclones are important agents of geomorphological change. They scour out river channels, deposit debris fans, cause slope failures, build up or break down coastal barriers, transform the nature of some coral islands (either building them up or erasing them), and change the turbidity and salinity of lagoons. Were their frequency, intensity and geogr[illegible] spread to change it would have significant implications. It is not, however, entirely clear just how much these important characteristics will change. Intuitively one would expect cyclone activity to become more frequent, intense and extensive if sea-surface temperatures were to rise, because SST is a clear control of where they develop. Indeed, there is a threshold at about 26.5–27.0 °C. However, the Intergovernmental Panel and some individual scientists are far from convinced that global warming will invariably stimulate cyclone activity.

Coral reefs may be sensitive to warming, partly because of the role that cyclones play in their evolution, partly because their growth can be retarded or accelerated because of changes in SSTs, and partly because their existence is so closely related to sea level.

In the 1980s there were widespread fears that if rates of sea-level rise were high (perhaps 2 to 3 m or more by 2100) then coral reefs would be unable to keep up and submergence of whole atolls might occur. Particular concern was expressed about the potential fate of Pacific Island groups, and of the Maldives in the Indian Ocean. However, with the reduced expectations for the degree of sea-level rise that may occur, there has arisen a belief that coral reefs may survive and even prosper with moderate rates of sea-level rise. As is the case with marshes and other wetlands, reefs are dynamic features that may be able to respond adequately to sea-level rise. It is also important to realize that their condition depends on factors other than the rate of submergence.

Increased sea-surface temperatures could have deleterious consequences for corals which are near their thermal maximum. Most coral species cannot tolerate temperatures greater than about 30 °C and even a rise in seawater temperature of 1–2 °C could adversely affect many shallow-water coral species. Increased temperatures in recent years have been identified as a cause of widespread coral bleaching (loss of symbiotic zooxanthellae). Those corals stressed by temperature or pollution might well find it more difficult to cope with rapidly rising sea levels than would healthy coral. Moreover, it is possible that increased ultraviolet radiation because of ozone layer depletion could aggravate bleaching and mortality caused by global warming. Various studies suggest that coral bl[illegible] the warm [illegible] 1980s [illegible] [illegible]nd Hayes 199[illegible]

How[illegible] Kinsey and Hopley (1991) believe that few of the reefs in the world are so close to the limits of temperature tolerance that they are likely to fail to adapt satisfactorily to an increase in ocean temperature of 1–2 °C, provided that there are not very many more short-term temperature deviations. Indeed, in general they believe that reef growth will be stimulated by the rising sea levels of a warmer world, and they predict that reef productivity could double in the next hundred years from around 900 to 1,800 million tonnes per year. They do, however, point to a range of subsidiary factors that could serve to diminish the increase in productivity: increased cloud cover in a warmer world could reduce calcification because of reduced rates of photosynthesis; increased rainfall levels and hurricane

activity could cause storm damage and freshwater kills; and a drop in seawater pH might adversely affect calcification.

However, reef accretion is not the sole response of reefs to sea-level rise, for reef tops are frequently surmounted by small islands (cays and motus) composed of clastic debris. Such islands might be very susceptible to sea-level rise. On the other hand, were warmer seas to produce more storms, then the deposition of large amounts of very coarse debris could in some circumstances lead to their enhanced development. However, the situation is complex, and in some cases potential vertical reef accretion could be reduced by storm attack. One also needs to consider changes in tropical storm frequency as well as changes in tropical storm magnitude, for high storm frequencies might change the relative importance of corals and calcareous algae (Spencer 1994).

Other coastlines

There are other coastlines that will also be substantially modified by sea-level rise resulting from global warming. These include sandy beaches, miscellaneous types of saltmarsh and areas of land subsidence.

Sandy beaches are held to be sensitive because of the so-called BRUUN RULE (Bruun 1962, see Plate 55). This predicts future rates of coastal erosion in response to rising sea level. Bruun envisaged a profile of equilibrium in which the volume of material removed during shoreline retreat is transferred onto the adjacent shoreface/inner shelf, thus maintaining the original bottom profile and nearshore shallow conditions. With a rise in sea level additional sediment has to be added to the below-water portion of the beach profile. One source of such material is beach erosion, and estimates of beach erosion of *c.*100 m for every 1 m rise in sea level have been postulated. However, although the concept is intuitively appealing, it is also difficult to confirm or quantify without precise bathymetric surveys and integration of complex nearshore profiles over a long period of time. Moreover, an appreciable time-lag may occur in shoreline response which is highly dependent upon local storm frequency. Furthermore, the model is essentially a two-dimensional one in which the role of longshore sediment movement is not considered. It is also assumed that no substantial offshore leakage of sediment occurs. Accurate determination of sediment budgets in three dimensions is still replete with problems. Whatever the problems of modelling, however, sandy beaches will tend to disappear from locations where they are already narrow and backed by high ground or swamp and marsh, but will probably tend to persist where they can retreat across wide beach ridge plains.

Saltmarshes, including MANGROVE SWAMPS, are potentially highly vulnerable to sea-level rise, particularly where sea defences and other barriers prevent the landward migration of marshes as sea-level rises. However, saltmarshes are dynamic features and in some situations may well be able to cope, even with quite rapid rises of sea level. Indeed, some important sediment trapping plants may extend their range in response to warming. Such plants include mangroves (e.g. in New Zealand) and also *Spartina anglica* (e.g. in northern Europe). They would tend to lead to an acceleration in marsh accretion.

One way of attempting to predict the effects of increasing rates of sea-level rise is to study those areas where the rates of sea-level rise are currently high because of subsidence. On the coast of south-east England, where rise occurs at a rate of 5 mm per year, saltmarshes appear to cope. Sediments eroded from the outer edge appear to contribute to the sediments which are accreted on the inner marsh surface. Moreover, UK saltmarshes have current rates of accretion that are the same order of magnitude as, or greater than, the predicted rates of sea-level rise.

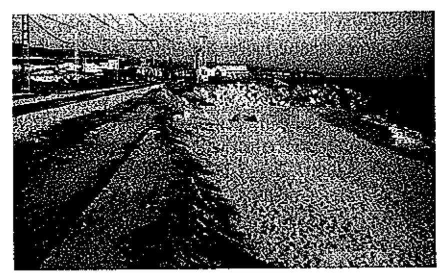


Plate 55 The main railway line between France and Spain near Barcelona. Note the severe erosion of the coastline, and the abandoned track in the foreground. Sandy beaches of this type will be especially sensitive to the effects of accelerated sea-level rise associated with global warming

Reed (1990) suggests that saltmarshes in riverine settings may receive sufficient inputs of sediment that they are able to accrete sufficiently rapidly to keep pace with projected rises of sea level. Likewise, some vegetation associations, e.g. *Spartina* swards, may be relatively more effective than others at encouraging accretion, and organic matter accumulation may itself be significant in promoting vertical build-up of some marsh surfaces. For marshes that are dependent upon inorganic sediment accretion, increased storm activity and beach erosion which might be associated with the greenhouse effect could conceivably mobilize sufficient sediments in coastal areas to increase their sediment supply.

One particular type of marsh that may be affected by anthropogenically accelerated sea-level rise is the mangrove swamp. Mangroves may respond rather differently to other marshes because their main plants are relatively long-lived trees and shrubs. This means that the speed of zonation change will be less. The degree of disruption is likely to be greatest in microtidal areas, where any rise in sea level represents a larger proportion of the total tidal range than in macrotidal areas. However, the setting of mangrove swamps will be very important in determining how they respond. River-dominated systems with large allochthonous sediment supply will have faster rates of shoreline progradation and deltaic plain accretion and so may be able to keep pace with relatively rapid rates of sea-level rise. By contrast, in reef settings in which sedimentation is primarily autochthonous, mangrove surfaces are less likely to be able to keep up with sea-level rises (Ellison and Stoddart 1990).

The ability of mangrove propagules to take root and become established in intertidal areas subjected to a higher mean sea level is in part dependent on species. In general the larger propagule species (e.g. *Rhizophora* spp) can become established in rather deeper water than can the smaller (e.g. *Avicennia* spp). The latter has aerial roots which project only vertically above tidal muds for short distances.

Mangrove colonization and migration would also be influenced by salinity conditions so that any speculations about mangrove response to sea-level rise must also incorporate allowance for change in rainfall and freshwater runoff.

In arid areas, such as the Middle East, great lengths of coastline are fringed by low level salt-plains (SABKHA). These features are generally regarded as equilibrium forms that are produced by depositional processes (e.g. alluvial siltation, aeolian inputs, evaporite formation, faecal pellet deposition) and planation processes (e.g. wind erosion and storm surge effects). They tend to occur at or about high tide level. Because of the range of depositional processes involved in their development they might be able to adjust to a rising sea level but quantitative data on present and past rates of accretion are sparse.

A crucial issue with all types of wetlands is the nature of the hinterland. Under natural conditions many marshes and swamps are backed by low-lying estuarine and alluvial land which could be displaced if a rising sea level were to drive the marshes landward. However, in many parts of the world sea defences, bunds and other structures have been built at the inner margins and these will prevent colonization of the hinterland. Experiments are now being conducted to see whether saltmarsh development can be promoted by the deliberate breaching of sea defences.

One final type of sensitive coastal environment is that where coastal submergence is taking place. The combination of local submergence with global sea-level rise will make these coasts especially prone to inundation. Some areas are subject to natural subsidence as a result of sediment loading on the crust (e.g. deltas) or because of tectonic processes, but some key areas are subjected to accelerated (i.e. anthropogenic) subsidence. This is brought about primarily by mining of ground water or hydrocarbons and can be especially serious in the case of coastal mega-cities, portions of which are either close to or beneath current sea level (e.g. Bangkok and Tokyo).

Although there may still be uncertainties about whether global warming will occur and about the various impacts of such warming should it occur, and although the degree of climatic and sea-level change that is being postulated might at first sight appear relatively modest, it would be wrong to be complacent about the potential geomorphological impacts brought about by global warming. Our knowledge of how geomorphological systems have reacted to the climatic fluctuations of the Holocene, and our knowledge of the intimate relationships between some geomorphological processes and climatic conditions, both lead us to the conclusion that some environments will respond in a manner that will be substantial in degree and which will have numerous consequences for human occupation of these environments.

References

Arnell, N. (1999) The impacts of climate change on water resources, in *Climate Change and its Impacts*, 14–17, Bracknell: UK Meteorological Office.

Bruun, P. (1962) Sea-level rise as a cause of shore erosion, *American Society of Civil Engineers Proceedings: Journal of Waterways and Harbors Division* 88, 117–130.

Elliot, J.G., Gellis, A.C. and Aby, S.C. (1999) Evolution of arroyos: incised channels of the southwestern United States, in S.E. Darby and A. Simon (eds) *Incised River Channels*, 153–185, Chichester: Wiley.

Ellison, J.C. and Stoddart, D.R. (1990) Mangrove ecosystem collapse during predicted sea level rise: Holocene analogues and implications, *Journal of Coastal Research* 7, 151–165.

Forman, S., Oglesby, R. and Webb, R.S. (2001) Temporal and spatial patterns of Holocene dune activity on the Great Plains of North America: megadroughts and climate links, *Global and Planetary Change* 29, 1–29.

Goreau, T.L. and Hayes, R.L. (1994) Coral bleaching and ocean 'hot spots', *Ambio* 23, 176–180.

Huybrechts, P., Litreguilly, A. and Reels, N. (1990) The Greenland ice sheets and greenhouse warming, *Palaeolgeography, Palaeoclimatology, Palaeoecology* 89, 399–412.

Intergovernmental Panel on Climate Change (2001) *Climate Change 2001: The Scientific Basis*, Cambridge: Cambridge University Press.

Kinsey, D.W. and Hopley, D. (1991) The significance of coral reefs as global carbon sinks – response to greenhouse, *Palaeogeography, Palaeoclimatology, Palaeoecology* 89, 363–377.

Koster, E.A. (1994) Global warming and periglacial landscapes, in N. Roberts (ed.) *The Changing Global Environment*, 150–172, Oxford: Blackwell.

Kwong, Y.T.J. and Tau, T.Y. (1994) Northward migration of permafrost along the Mackenzie Highway and climatic warming, *Climatic Change* 26, 399–419.

Muhs, D.R. and Maat, P.B. (1993) The potential response of aeolian sands to greenhouse warming and precipitation reduction on the Great Plains of the United States, *Journal of Arid Environments* 25, 351–361.

Reed, D.J. (1990) The impact of sea level rise on coastal saltmarshes, *Progress in Physical Geography* 14, 465–481.

Shiklomanov, I.A. (1999) Climate change, hydrology and water resources: the work of the IPCC, 1988–1994, in J.C. van Dam (ed.) *Impacts of Climate Change and Climate Variability on Hydrological Regimes*, 8–20, Cambridge: Cambridge University Press.

Spencer, T. (1994) Tropical coral islands – an uncertain future, in N. Roberts (ed.) *The Changing Global Environment*, 190–209, Oxford: Blackwell.

Woo, M.K., Lewkowicz, A.G. and Rouse, W.R. (1992) Response of the Canadian permafrost environment to climate change, *Physical Geography* 13, 287–317.

A.S. GOUDIE

GOLDICH WEATHERING SERIES

The types and proportions of various minerals in a weathering profile are usually quite different from the original bedrock. Some minerals seem to survive more or less unaltered even after being subjected to prolonged WEATHERING, while others decompose more rapidly. In many weathering studies, the silicate minerals, the primary constituents of igneous and metamorphic rocks, are arranged into an order of susceptibility to chemical weathering. The most commonly cited order was first proposed by S.S. Goldich (1938), based on a detailed study of the mineralogic changes of granitoid rocks during weathering (Figure 78).

Goldich (1938) concluded that minerals which form at high temperatures and pressures (olivine, amphiboles, pyroxenes, calcium plagioclase), and hence are the first to precipitate, are markedly less stable and weather much more quickly than minerals which crystallize at lower temperatures and pressures (sodium plagioclase, potassium feldspar, micas and quartz) (Figure 78). This sequence is the reverse of BOWEN'S REACTION SERIES, which ranks minerals in their order of crystallization from a melt.

CHEMICAL WEATHERING reactions are with the cations that bind the silica structural units together. Thus the relative strength between the oxygen and cations in each mineral and the structure of the bonding are both significant. The isolated Si-O tetrahedra in olivine are the least stable in weathering; while quartz, which is completely formed of interlocking silica tetrahedra with no intervening cations, is the most stable. If muscovite and the plagioclases are disregarded, the order of the Goldich weathering series coincides with the classification of silicate structures, based on increasing Si–O–Si bonds from zero in olivine to four in quartz.

Although Goldich's series has widespread applications and usually works well, local exceptions have been documented.

Reference

Goldich, S.S. (1938) A study of rock weathering, *Journal of Geology* 46, 17–58.

SEE ALSO: Bowen's reaction series; chemical weathering

CATHERINE SOUCH

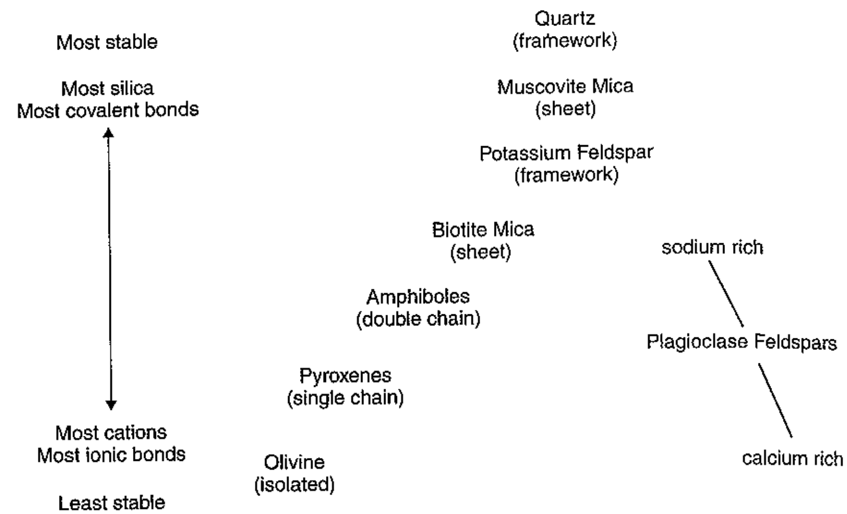


Figure 78 Goldich weathering series

GORGE AND RAVINE

Gorges, which may be hundreds of metres deep, are caused either by incision of a river against an uplifting landmass, the superimposition of a channel across resistant rock, the outburst of floodwaters across a landscape, or by the headward retreat of a KNICKPOINT or WATERFALL (Rashleigh 1935; Derricourt 1976; Tinkler *et al.* 1994; van der Beek *et al.* 2001). Ravines are much smaller gashes (the order of metres to tens of metres wide and deep) cut into the weak bedrock, or frequently into superficial sediments such as glacial deposits or deeply weathered horizons. The term ravine is frequently used in the context of soil erosion and land degradation, and a ravine, or ravine network, will have steep, weakly consolidated side slopes, flat channel bottoms characterized by a heavy sediment load, and a clear break of slope with the surface above. Present academic literature seems to find the technical use of the word limited to south and east Asia (Raj *et al.* 1999), otherwise it is used as a synonym for gully. It may occur as the generic part of a place name: Yamuna Ravine in India, Elk Ravine, New Hampshire, USA.

Neither of the terms *gorge* nor *ravine* is well defined in the literature, although a Gorge (e.g. the Three Rivers Gorge in western China) is generally understood to be typical of rivers of larger sizes. The word 'ravine' tends to imply a small deeply incised channel in a low-order drainage basin. Both terms imply a river deeply incised below the surrounding landscape, local slope processes being unable to reduce the side slopes at the same rate that the river is incising into the terrain. Thus there is often little sensitivity to the local topography. Large, deep gorges require mechanically strong country rock, although the typically steep valley slopes may still be susceptible to failure by rock fall and rockslides. Gorges are frequently found in areas where drainage is antecedent upon actively growing fold systems such as the Himalayan ranges, or where it is superimposed (superposed) upon more resistant rocks from weaker cover rocks.

In exceptionally large river systems the term gorge usually refers to the deeply incised, and often scarcely accessible inner gorge (Kelsey 1988). Bedrock channels often contain an inner channel, where bedload transport rates are

highest, and erosion is enhanced. Because of the large variations in discharge which have occurred in many high latitude drainage basins during the Quaternary, it is unclear to what extent the inner channel of a river system carrying large glacial outflow discharges becomes the inner gorge of its non-glacial successor. Excavations for the Boulder Dam on the Black Canyon of the Colorado revealed an unsuspected inner gorge up to 25 m below flanking bedrock edges to the channel (Legget 1939: 322–323).

Catastrophic scale outburst floods of glacial stored waters are another mechanism, unsuspected until recent decades, for the formation of gorges (Baker 1978; O'Connor 1993; Rathburn 1993; Knudsen *et al.* 2001). Such floods may have been repeated many times during the Quaternary, their cumulative sum affect being what we now see.

Scheidegger *et al.* (1994) argue for strong structural control by large-scale regional joints and fault systems, in the geographical layout of large gorges. However overall trends in gorge orientation usually owe their origin to regional scale topographic trends (Baker 1978; Rathburn *et al.* 1993), to which structural control merely adds local detail. Subsequent river erosion may generate entrenched or incised meandering patterns unrelated to local or regional structure.

Buried gorges are not uncommon in glaciated terrains, many being found in the Great Lakes region of eastern Canada (Davis 1884; Karrow and Terasmae 1970; Greenhouse and Karrow 1994). The infill of permeable glacial sediments within a bedrock gorge often produces localities favourable for groundwater exploration (Farvolden 1969).

References

Baker, V.R. (1978) Paleohydraulics and hydrodynamics of Scabland Floods, in V.R. Baker and D. Nummedal (eds) *The Channeled Scabland*, 59–80, Washington, DC: NASA.

Davis, W.M. (1884) Gorges and waterfalls, *American Journal of Science* 28, 123–132.

Derricourt, R.M. (1976) Retrogression rate of the Victoria Falls and the Batoka Gorge, *Nature* 264, 23–25.

Farvolden, R.N. (1969) Bedrock channels of southern Alberta, in J.G. Nelson and M.J. Chambers (eds) *Geomorphology: Process and Methods in Canadian Geography*, 243–255, Toronto: Methuen.

Greenhouse, J.P. and Karrow, P.F. (1994) Geological and geophysical studies of buried valleys and their fills near Elora and Rockwood, Ontario, *Canadian Journal of Earth Sciences* 31, 1,838–1,848.

Karrow, P.F. and Terasmae, J. (1970) Pollen-bearing sediments of the St. Davids buried valley at the Whirlpool, Niagara Gorge, Ontario, *Canadian Journal of Earth Sciences* 7, 539–542.

Kelsey, H.M. (1988) The formation of inner gorges, *Catena* 15, 433–458.

Knudsen, K.L., Sowers, J.M., Ostenaa, D.A. and Levish, D.R. (2001) Evaluation of glacial outburst flood hypothesis for the Big Lost River, Idaho, *Ancient Floods, Modern Hazards*, Washington, DC: American Geophysical Union.

Legget, R.E. (1939) *Geology and Engineering*, New York: McGraw-Hill.

O'Connor, J.E. (1993) Hydrology, hydraulics, and geomorphology of the Bonneville Flood, *Geological Society of America Special Paper* 274.

Raj, R., Maurya, D.M. and Chamyal, L.S. (1999) Tectonic control on distribution and evolution of ravines in the lower Mahi Valley, Gujarat, *Journal of the Geological Society of India* 53(6), 669–674.

Rashleigh, E.C. (1935) *Among the Waterfalls of the World*, London: Jarrolds.

Rathburn, S.L. (1993) Pleistocene cataclysmic flooding along the Big Lost River, east central Idaho, *Geomorphology* 8, 305–319.

Scheidegger, A.E. (1994) On the genesis of river gorges, *Transactions, Japanese Geomorphological Union* 15(2), 91–110.

Tinkler, K.J., Pengelly, J.W., Parkins, W.G. and Asselin, G. (1994) Postglacial recession of Niagara Falls in relation to the Great Lakes, *Quaternary Research* 42, 20–29.

Van der Beek, P., Pulford, A. and Braun, J. (2001) Cenozoic landscape development in the Blue Mountains (SE Australia): lithological and tectonic controls on Rifted Margin Morphology, *Journal of Geology* 109, 35–56.

KEITH J. TINKLER

GPS

The Global Positioning System (GPS) is a constellation of satellites developed by the US Department of Defense to provide precise positioning and navigation information. GPS receivers determine position through repeated measurements of digitally tagged radio signals from the satellites. Conceived for military purposes, the commercial applications for positioning information have blossomed. Analysts have suggested that the global GPS market is worth over US$16 billion (in 2002). Among the varied users of GPS are geomorphologists requiring geo-referenced positioning information for field terrain. However, the wide variety of systems available and the enormous range in cost and capability requires caution on the part of the user and an ability to assimilate a multitude of jargon and proprietary software.

Initially, the Department of Defense used a procedure termed Selective Availability to dither the precise time code and degrade the accuracy of the signal. This has been set to zero since May 2000, improving reliability and consistency though applications requiring sub-metre accuracy (and hence all fieldwork requiring elevation data) continue to require differential GPS (DGPS). DGPS relies on a static reference receiver at a known control point which logs bias errors over the same time period that another receiver (the 'rover') is occupying the points of interest. The measured errors are used to correct the rover position either by downloading and 'post-processing' the data, or by receiving corrections via radio telemetry (known as 'real time kinematic'). The control point can be operated by the user or be a commercial ground station broadcasting corrections. The reference frame for GPS output is the World Geodetic System 1984 (WGS-84), a geocentric system returning ellipsoid co-ordinates in latitude and longitude. Altitude is derived as elevation above the ellipsoid and some knowledge is needed to integrate GPS-derived height data with existing levelling data or to translate positions into a local datum. Fortunately there are many textbooks providing technical details (e.g. Hofmann-Wellenhof *et al.* 2001). Relatively few papers consider explicitly GPS applications in geomorphology (Cornelius *et al.* 1994; Fix and Burt 1995; Higgitt and Warburton 1999) but an increasing number make routine use of GPS as part of the data-gathering procedure. Four broad areas of application can be identified:

Rectification

Global referencing is essential in most geomorphological research. GPS can assist observations where detailed maps are lacking or it can be used for registering ground markers to analyse aerial photographs or remotely sensed imagery. This is useful for assessing change in sequential imagery such as the dynamics of land degradation (Gillieson *et al.* 1994). In terrain remote from conventional benchmarks, GPS can save much time in establishing the elevation of sample points.

Detailed topographic survey

The speed of GPS data capture offers scope for producing accurate digital elevation models (DEMs) of moderately sized field areas. The abundance of points in a GPS survey generates topographic attributes which can be used as input in hydrological models. A related commercial development is 'precision agriculture' where GPS receivers mounted to farm vehicles produce detailed information about spatial variations in crop yields or soil conditions. One consequence of precision, as highlighted by Wilson *et al.* (1998), is the recognition that calculated topographic attributes are sensitive to the resolution and distribution of survey points. By implication, estimation of topographic variables from a limited number of survey points may be prone to large errors. A dense network of GPS survey points around a catchment can provide a more enlightened summary about the statistical distribution of slope characteristics.

Measuring change in landforms

GPS is ideal for measuring sequential change in landform characteristics. Geologists have made extensive use of networks of high precision GPS for identifying ground movements associated with earthquakes and volcanic eruptions. Geomorphological applications are apparent in neotectonics and landslide research. Where budgets are more restrictive, repeat surveys provide similar information. This has been used to construct detailed maps of river channel change (Brasington *et al.* 2000). As the object of geomorphological study is usually inanimate, there are no parallels to ecological applications that examine animal behaviour. The methodology to determine grazing patterns by fitting ungulates with GPS collars might be adapted to keep track of students during field trips!

Geomorphological mapping

Where acquisition of elevation data are not critical, GPS can be an effective mapping tool. The outline of geomorphological features (e.g. the edge of river terraces or landslides) or point patterns (e.g. glacial erratics) can be obtained speedily in terrain where conventional surveying is impractical and the features cannot be determined sufficiently from aerial photography. GPS software has a facility to tag attribute information to the data and can be integrated into GIS. It should be remembered that the GPS receiver requires an unobstructed path to the satellites and hence mapping in mountainous terrain, urban environments or under forest cover can be problematic.

In each of the categories above, the speed and frequency of positioning is the essential difference enabling GPS to provide data that would be difficult or impossible to derive from conventional surveying methods. As such, GPS is not a technique producing completely new data but rather an application that improves accuracy and/or frequency of measurement coupled with efficient data-processing capability. The cost of GPS receivers spans at least two orders of magnitude. High precision GPS is not only expensive but requires a thorough understanding of surveying principles and the equipment can be bulky. Mapping grade GPS is highly portable and can be operated by a single user where safety considerations allow. The required accuracy should dictate the specification of GPS but its subsequent application in geomorphology is wide-ranging.

References

Brasington, J., Rumsby, B.T. and McVey, R.A. (2000) Monitoring and modelling morphological change in a braided gravel-bed river using high resolution GPS-based survey, *Earth Surface Processes and Landforms* 25, 973–990.

Cornelius, S.C., Sear, D.A. and Carver, S.J. (1994) GPS, GIS and geomorphological field work, *Earth Surface Processes and Landforms* 19, 777–787.

Fix, R.E. and Burt, T.P. (1995) Global Positioning System: an effective way to map a small area or catchment, *Earth Surface Processes and Landforms* 20, 817–828.

Gillieson, D.S., Cochrane, J.A. and Murray, A. (1994) Surface hydrology and soil movement in an arid karst – the Nullabor Plain, Australia, *Environmental Geology* 23, 125–133.

Higgitt, D.L. and Warburton, J. (1999) Applications of differential GPS in upland fluvial geomorphology, *Geomorphology* 31, 411–439.

Hofmann-Wellenhof, B., Lichtenegger, H. and Collins, J. (2001) *GPS: Theory and Practice*, 5th edition, Heidelberg: Springer-Verlag.

Wilson, J.P., Spangrud, D.J., Nielsen, G.A., Jacobsen, J.S. and Tyler, D.A. (1998) Global positioning system sampling intensity and pattern effects on computed topographic attributes, *Soil Science Society of America Journal* 62, 1,410–1,417.

DAVID HIGGITT

GRADE, CONCEPT OF

Since the end of the seventeenth century (Dury 1966; Chorley 2000), various engineers have been concerned both with the regulation of natural rivers and with the operation and construction of artificial channels. This required an interest in geometric stability or equilibrium (grade) which tended to be roughly constant or subjected to limited oscillation over a recognized period of time. Such stability could arise from some sort of balance between, for example, fluid shear stress and material resistance, or some equalization between those sedimentary processes (e.g. cut-and-fill) which control channel morphology. The concept entered mainstream geomorphology through G.K. Gilbert (1877), for whom the major geometrical evidence of the graded state was a smooth, concave-up river long profile. Grade was also accommodated within the Davisian cycle, and for Davis (1902) the elimination of breaks of slope was the hallmark of the graded condition. A major contribution to understanding the concept of grade was made by Mackin (1948) who defined a graded river as: 'one in which, over a period of years, slope and channel characteristics are delicately adjusted to provide, with available discharge... just the velocity required for the transportation of the load supplied from the drainage basin'.

In 1965 Schumm and Lichty introduced the concept of a time-span intermediate between the longer interval of 'cyclic time' and the shorter period of 'steady time'. They defined graded time as 'a short span of cyclic time during which a graded condition or dynamic equilibrium exists'. Later, Schumm (1977) saw a graded stream as 'a process-response system in steady-state equilibrium, and the equilibrium is maintained by self-regulation or negative feedback, which operates to counteract or reduce the effects of external change on the system so that it returns to an equilibrium condition'.

References

Chorley, R.J. (2000) Classics in physical geography revisited, *Progress in Physical Geography* 24, 563–578.

Davis, W.M. (1902) Base level, grade and peneplain, *Journal of Geology* 10, 77–111.

Dury, G.H. (1966) The concept of grade, in G.H. Dury (ed.) *Essays in Geomorphology*, 211–233, London: Heinemann.

Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, Washington, DC: US Geological Survey.
Mackin, J.H. (1948) Concept of the graded river, *Geological Society of America Bulletin* 59, 463–512.
Schumm, S.A. (1977) *The Fluvial System*, New York: Wiley.
Schumm, S.A. and Lichty, R.W. (1965) Time, space and causality in geomorphology, *American Journal of Science* 263, 110–119.

A.S. GOUDIE

GRADED TIME

The most concise description of graded time is derived from Mackin (1948):

> A graded river is one in which, over a period of years, slope and channel characteristics are delicately adjusted to provide, with available discharge, just the velocity required for the transportation of the load supplied from the drainage basin. The graded stream is a system in equilibrium; its diagnostic characteristic is that any change in any of the controlling factors will cause a displacement of the equilibrium in a direction that will tend to absorb the effect of the change.

It is clear that 'graded time' is the time over which a stream is in balance in this way.

A second, less useful, term is the 'time to grade' (see GRADE, CONCEPT OF) or the time that it takes for a river to attain a graded condition. This is not a simple idea because the graded condition is not reached throughout all parts of a system at the same time. In rivers, for example, W.M. Davis (1902) stated that grade would be attained first in the lower reaches and then extend upstream. It would also be attained first in the most adjustable materials.

The term, graded time, therefore has a spatial dimension. Davis (1899) stated that when the trunk streams were graded the stage of early maturity had been reached, when the smaller headwaters were graded maturity was well advanced and when even the wet river rills and the waste mantle were graded the stage of old age had been attained.

The idea that once grade had been achieved the balance of forces, sediment loads and forms would remain adjusted even though the landscape was still being slowly lowered has always been an uncomfortable element of the geographical cycle. The timeless aspect of the concept of grade does not sit happily within the timebound cyclical framework.

References

Davis, W.M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.
——(1902) Base level, grade and peneplain, *Journal of Geology* 10, 77–111.
Mackin, J.H. (1948) Concept of the graded river, *Geological Society of America Bulletin* 59, 463–512.

DENYS BRUNSDEN

GRANITE GEOMORPHOLOGY

Granite terrains of the world, whether in lowland, upland or mountain settings, often have distinctive morphology, different from one typical for the surrounding country rock. Although it would probably be impossible to find a landform endemic for granite, many are most prominent if bedrock is granitic. Examples include boulders, TORS, INSELBERGS, BORNHARDTS, INTERMONTANE BASINS, and a range of microforms such as WEATHERING PITS or TAFONI (Twidale 1982). They usually form through selective bedrock weathering, either in subsurface (see DEEP WEATHERING) or at the topographic surface, followed by evacuation of the loose products of rock disintegration. However, there is no 'standard' granite landscape, as these can be significantly different, even if located adjacent to each other. Granite is known to support extensive plains of extreme flatness and, by contrast, high-mountain, highly dissected terrains. In spite of widespread presence of a weathering mantle, bedrock frequently crops out at the topographic surface and tors and boulder fields are characteristic landmarks. Granite is typically, but by no means universally, more resistant to weathering and erosion than surrounding country rock, and therefore tends to form upland terrains and to support topographic steps.

Lithological and structural properties of granite, such as mineral composition, texture and joint density, which are often highly variable within a single granite intrusion, are the keys to understanding the selectivity of weathering and the prominence of many small- and medium-scale granite landforms.

Granites are usually fairly regularly jointed according to an orthogonal pattern, i.e. they are

cut by three subsets of joints perpendicular to themselves, which delimit cuboid block compartments. As fractures guide movement of ground water through the rock mass, weathering acts most efficiently along joints and preferentially attacks the sides and edges of joint-bound cubes, which results in their progressive rounding and the typical multi-convex appearance of many granite landscapes. In the subsurface, weathering attack along joints transforms sharp-edged blocks into rounded core stones surrounded by a thoroughly disintegrated mass. Furthermore, because of variable joint density over short distances (< 10 m) significant differences in the intensity of rock disintegration may occur. Less fractured parts are left standing as rock pillars or castellated tors, whereas adjacent more closely jointed compartments are disintegrated into block rubble or GRUS. Evacuation of weathered material reveals a range of topographically negative features, common for granite areas. These include rock basins developed either at joint intersections or between master joints, and linear joint-guided valleys.

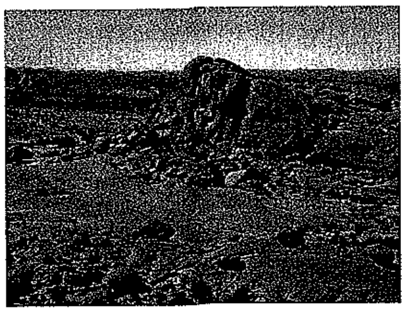
Many post-orogenic granites are typically very massive, with large-scale SHEETING joints being dominant. Joint spacing in such granites can be extremely wide, more than 10 m apart. In these areas topography usually follows the curvature of sheeting planes, bornhardts are common, and minor weathering features on rock surfaces often grow to gigantic dimensions.

Rock texture is equally important. Coarse variants of granite with abundant large phenocrysts of K-feldspar usually support a varied, rough relief, with big boulders, inselbergs, and intervening basins. The majority of domed inselbergs and bornhardts seems to be built of massive, coarse-grained granite. Likewise, minor features on rock surfaces are best developed within coarse granite. Finer variants tend to give rise to a more subdued topography, often with frequent angular tors.

Another factor important for the development of granite topography is mineralogical and chemical composition of the rock, including proportions between quartz and feldspar, between different types of feldspar, silica content, and proportions between potassium, sodium and calcium. Potassium-rich granites tend to be more resistant and therefore often form higher ground and give rise to spectacular inselberg landscapes, whereas granites with high plagioclase content typically underlie gently rolling terrains and low ground (Brook 1978; Pye *et al.* 1986).

In areas, where high precipitation and humidity levels favour deep weathering, subsurface decomposition of granite becomes crucial in the evolution of topography (Twidale 1982). Granite terrains, except for those in arid areas or in high mountains, usually carry a spatially extensive, thick mantle of weathering residuals. A very wide range of thicknesses has been reported, from only a few to as much as 200–300 m (Ollier 1984). There are different types of weathering mantles developing on granite, but rather shallow grus and more advanced geochemically, kaolinite-rich covers are most typical. This division likely reflects environmental conditions during weathering, including climatic conditions, their change through time and geomorphic stability of the surface. What both categories of granite weathering mantles have in common though, is the rough topography of the weathering mantle/bedrock interface (i.e. WEATHERING FRONT), attributable to the selectivity of deep weathering, and the usually sharp nature of this boundary. Therefore, stripping of the pre-weathered material often reveals complicated bedrock topography, with numerous low domes, isolated boulders and tor-like bedrock projections separated by basins and linear hollows. Indeed, many granite landscapes are interpreted to be the product of two-stage development, with the phase, or phases, of deep selective weathering followed by stripping and exposure of weathering front topography (Plate 56). The presence of bornhardts, tors and rounded boulders is occasionally used to infer the two-stage evolution, even if no remnants of any weathering mantle are left and no independent evidence exists that such ever existed. However, examples from areas with a long history of aridity such as the Namib Desert demonstrate that deep weathering is not a necessary precursor to the development of multi-convex granite topography, which primarily reflects structural control (Selby 1982).

The majority of detailed studies has concentrated on prominent medium-scale landforms such as boulder fields, tors, bornhardts and inselbergs, and pediments, or distinctive minor features of rock surfaces. Analyses of entire landform assemblages and their evolution through time are fewer. One of the attempts has been made by Thomas (1974) who distinguished multi-concave, multi-convex and stepped or

(a)


(b)


Plate 56 Granite landscapes share many of their characteristics regardless of the climatic zone in which they occur. Both the granite landscape of (a) the Erongo massif in arid Namibia and (b) the humid Estrela Mountains in central Portugal are dominated by massive domes, big rounded boulders scattered around and basins formed through selective joint-guided weathering

multistorey landscapes. In multi-concave terrains, topographic basins of various sizes are dominant features. Their occurrence may be related to either inliers of less resistant granite or to the occurrence of initially more jointed rock compartments (Thorp 1967; Johansson *et al.* 2001). Multi-convex terrains are those typified by closely spaced domes, or similar upstanding rock masses, so there is little space left for basins to develop. They are common in homogeneous, poorly jointed intrusions, where lines of structural weakness available for exploitation by weathering are few. Another type of multi-convex landscape is one dominated by low hills weathered throughout, possibly with a solid rock core.

Stepped landscapes are characterized by the presence of topographic scarps separating successive levels or 'storeys'. They typically occur in areas subjected to recent, but moderate uplift which was proceeding concurrently with weathering and stripping. Since the scarps are apparently not tectonically controlled, it is proposed that they form due to reduced rates of advance of the weathering front at progressively higher topographic levels, whereas their exact location reflects the occurrence of a more massive granite (Wahrhaftig 1965; Bremer 1993). In each of these landscape types, spatial patterns of individual landforms are largely controlled by lithology and structure.

Specific landform assemblages typify ring complexes, made of concentrically arranged intrusions of granite and other rocks, differing in mineralogy and texture, and intersected by dykes. Depending on the susceptibility of particular complex-forming rock units to weathering and erosion, a concentric pattern of uplands alternating with basins develops. Most resistant dykes form linear ridges, sculpted into jagged rock crests.

In addition, granite landscapes may take the form of a plain, either rock-cut or deeply weathered, as it is common in Australia, parts of Africa, or Scandinavia. Within strongly uplifted and highly dissected areas, a mountainous all-slope topography evolves (Twidale 1982). In both cases, structural control is less obvious and its influence surpassed by the high efficacy of planation or dissection.

Granite geomorphology has played an important part in CLIMATIC GEOMORPHOLOGY, and especially in the attempts to use specific landforms as indicators of specific climatic conditions. For instance, claims have been made that granite domes evolve in the humid tropics, boulder heaps are more typical for seasonally dry areas, whereas small-scale flutings indicate hot and humid conditions (Wilhelmy 1958). Moreover, the apparent durability of granite and its ability to withstand high compressive and tensile stresses have been used to support the claim that granite landforms, once formed under distinctive environmental conditions, may survive many subsequent environmental changes. In some Central European studies, minor granite landforms have been used to establish the chronology of denudation and environmental change since the mid-Tertiary. Increasing recognition of pervasive structural and lithological control on the evolution of granite landforms, as well as of the crucial role of subsurface weathering, have seriously undermined the basis of the climatic approach to granite geomorphology. At present, a consensus appears to have been reached that the evolution and appearance of granite landscapes are primarily controlled by structure, and similarities of structures explain why granite landform assemblages in contrasting geographical settings often look very much the same.

Many of the geomorphologically classic landforms and landscapes are underlain by granite. Examples include the tors of Dartmoor in south-west England, domes and U-shaped glacial valleys of the Yosemite National Park in Sierra Nevada, USA, sugar-loaf hills in Rio de Janeiro, African inselberg landscapes of Nigeria, Kenya and Namibia, fluted coastal outcrops in the Seychelles, and the Wave Rock in Western Australia.

References

Bremer, H. (1993) Etchplanation, review and comments of Büdel's model, *Zeitschrift für Geomorphologie N.F., Supplementband* 92, 189–200.
Brook, G.A. (1978) A new approach to the study of inselberg landscapes, *Zeitschrift für Geomorphologie N.F., Supplementband* 31, 138–160.
Johansson, M., Migoń, P. and Olvmo, M. (2001) Joint-controlled basin development in Bohus granite, SW Sweden, *Geomorphology* 40, 145–161.
Ollier, C.D. (1984) *Weathering*, London: Longman.
Pye, K., Goudie, A.S. and Watson, A. (1986) Petrological influence on differential weathering and inselberg development in the Kora area of Central Kenya, *Earth Surface Processes and Landforms* 11, 41–52.
Selby, M.J. (1982) Form and origin of some bornhardts of the Namib Desert, *Zeitschrift für Geomorphologie N.F., Supplementband* 26, 1–15.
Thomas, M.F. (1974) Granite landforms: a review of some recurrent problems of interpretation, in *Institute of British Geographers, Special Publication* 7, 13–37.
Thorp, M. (1967) Closed basins in Younger Granite Massifs, northern Nigeria, *Zeitschrift für Geomorphologie N.F., Supplementband* 11, 459–480.
Twidale, C.R. (1982) *Granite Landforms*, Amsterdam: Elsevier.
Wahrhaftig, C. (1965) Stepped topography of the southern Sierra Nevada, *Geological Society of America Bulletin* 76, 1,165–1,190.
Wilhelmy, H. (1958) *Klimamorphologie der Massengesteine*, Braunschweig: Westermann.

Further reading

Gerrard, J. (1986) *Rocks and Landforms*, London: Unwin Hyman.
Godard A., Lagasquie, J.-J. and Lageat, Y. (2001) *Basement Regions*, Berlin: Springer.
Lageat, Y. and Robb, L.J. (1984) The relationships between structural landforms, erosion surfaces and the geology of the Archaean granite basement in the Barberton region, Eastern Transvaal, *Transactions Geological Society of South Africa* 87, 141–159.
Twidale, C.R. (1993) The research frontier and beyond: granitic terrains, *Geomorphology* 7, 187–223.

PIOTR MIGOŃ

GRANULAR DISINTEGRATION

Granular disintegration is the physical disintegration of rock into individual grains and rock crystals. The product of granular disintegration is usually coarse-grained, loose debris, which can be easily removed by erosive agents such as wind, water and gravity. This form of rock breakdown occurs commonly in coarse-grained rocks such as sandstone, dolerite and granite. Clay-rich rocks are thought to be particularly susceptible (Smith *et al.* 1994).

The surface grains and rock crystals which become detached may be unweathered and unaltered. They may also remain *in situ* but could be easily removed by light brushing with the hand. Where the product of granular disintegration remains *in situ* and accumulates, a gritty SAPROLITE is produced, known as GRUS. When loose material is removed, the fresh surface beneath may be pitted and uneven. The loose material may accumulate as a sandy deposit.

There are a number of mechanical and chemical mechanisms of granular disintegration and it is likely that the process can be attributed to several or all of these. It is equally likely that more than one of the mechanisms operates simultaneously in many cases:

1 *Solution of soluble cement* Sandstones cemented by soluble calcareous material are particularly susceptible to granular disintegration due to this mechanism.
2 *Stress induced by volumetric expansion* The growth of salt and ice crystals leads to a volumetric expansion. Under certain conditions, this can produce sufficient force to rupture the rock and this is most likely to occur at locations of weakness such as grain boundaries. There is ample evidence that rocks readily disintegrate in salt-rich environments due to salt crystallization (Evans 1970). Experimental work has also shown salt to be extremely effective in the

physical breakdown of rock (e.g. Goudie *et al.* 1970). Chemical weathering processes such as hydrolysis may involve expansion of minerals sufficient to produce crystal fracture.

3 *Release of residual stress* This is stress in rock due to primary crystallization or lithification. These stresses exist in a balanced state in unweathered rock. However, residual stresses can become unbalanced, and therefore released, by erosion, weathering and mass movement. The stresses generated can be large enough to cause crack propagation (e.g. Bock 1979).

4 *Water adsorption* Repeated wetting and drying may be responsible for the disintegration of fine-grained rocks such as mudstone (see SLAKING). Water molecules are absorbed onto mineral surfaces and may produce force sufficient to prise particles apart.

References

Bock, H. (1979) A simple failure criterion for rough joints and compound shear surfaces, *Engineering Geology* 14, 241–254.
Evans, I.S. (1970) Salt crystallisation and rock weathering: a review, *Revue de Géomorphologie Dynamique* 19, 153–177.
Goudie, A.S., Cooke, R.U. and Evans, I.S. (1970) Experimental investigation of rock weathering by salts, *Area* 2, 42–48.
Smith, B.J., Magee, R.W. and Whalley, W.B. (1994) Breakdown patterns of quartz sandstone in a polluted urban environment, Belfast, Northern Ireland, in D.A. Robinson and R.B.G. Williams (eds) *Rock Weathering and Landform Evolution*, 131–150, Chichester: Wiley.

Further reading

Cooke, R.U. (1981) Salt weathering in deserts, *Proceedings of the Geologists' Association* 92, 1–16.
Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

DAWN T. NICHOLSON

GRAVEL-BED RIVER

An alluvial river in which the average diameter of bed materials exceeds 2 mm. An upper grain-size limit is seldom identified, but channels with beds predominantly composed of very large, essentially immobile boulders (>256 mm) may be regarded as a distinct type or subcategory, especially if they exhibit step-pool morphology (see STEP-POOL SYSTEM). The primary distinction is with SAND-BED RIVERS (bed material 0.063–2 mm). Gravel-bed rivers dominate in upland and piedmont settings where the sediment supplied to the channel is coarse and poorly sorted. With distance downstream, bed materials become smaller (see DOWNSTREAM FINING) and an abrupt gravel–sand transition often terminates the gravel reach.

Although gravel-bed rivers transport significant quantities of sand, much of it in suspension, the proportion of BEDLOAD transport is apt to be higher than in sand-bed channels. Parker (in press) usefully defines a limit case wherein the median bed-particle size is greater than 25 mm and bedload transport dominates. Examination of such channels reveals that the boundary shear stresses generated by modest flows (for example, bankfull) are barely capable of moving median grain-sizes. In sharp contrast to sand-beds, gravel-bed channels are therefore characterized by hydraulic stresses that rarely exceed the entrainment thresholds of particles exposed at the bed surface, and large floods are required to generate significant sediment transport. Sediment yield is limited by the competence of flows to move the coarse load, rather than the availability of mobile sediments per se. This is a definitive characteristic of gravel-bed rivers, though in many environments vertical sorting of the bed material (ARMOURING) does significantly limit the availability of potentially mobile, subsurface sediments.

Close to the threshold for motion, the coarse armour layer remains intact and transport involves individual grain movements across its largely unbroken surface. Once rotated out of bed pockets by lift and drag forces, particles roll and bounce across the bed, intermittently stopping in stable positions from where they may be entrained again by instantaneous turbulent stresses. Particles cover relatively short distances, potentially falling into stable interstices or pockets in the armour layer. This marginal transport regime dominates during most floods, with the armour layer moderating sediment supply and grain velocities. However, as flow intensity increases, larger areas of the armour layer are breached, the number of particles in motion rises and, during exceptional floods, most of the bed may be mobile. Even during mass transport, flows are seldom sufficiently deep to form mobile bedforms of the geometry found in sand-bed channels (steep ripples and dunes), but low-amplitude forms known as gravel sheets are common, and their passage generates bedload pulses.

A number of small-scale bedforms are recognized in gravel-bed rivers and are important because they influence near-bed hydraulics and, like bed armour, moderate sediment supply and entrainment. Pebble clusters that form when large obstacle clasts distort the flow and impede the passage of other clasts, are repeating, streamlined features. Transverse ribs are regularly spaced, linear ridges of coarse clasts that form perpendicular to the flow under supercritical conditions. Stone cells are reticulate structures that may form where transverse ribs and pebble clusters intersect.

These micro-bedforms protrude above the bed surface and therefore contribute to overall flow resistance, as do channel-scale grain accumulations (bars and riffles) that retard the passage of water. However, in contrast to sand-bed channels where bedforms dominate boundary resistance, grain roughness is regarded as the dominant component in gravel-bed rivers (see ROUGHNESS).

The longitudinal profiles of gravel-bed rivers typically exhibit significant concavity that reflects adjustment to downstream fining and the associated reduction in competence required to transport a given load. Channel gradients therefore vary significantly from as much as 0.1 to as little as 0.001.

Cross-sectional form is determined by numerous variables in addition to bed-material size. Nevertheless, for a given discharge gravel-bed rivers do tend to be shallower than sand-bed channels and have higher width–depth ratios. This reflects the dominance of bedload transport and a lack of fine-grained, floodplain deposition, that together promote lateral instability and channel widening. Local variability of width and depth is exacerbated in gravel-bed rivers by well-developed riffle-pool sequences. Analogous bed topography is evident in some sand-bed channels, bedrock channels, and as step-pools in boulder-bed channels, but riffle-pools are best developed in gravelly channels with heterogeneous bed materials.

Gravel-bed rivers may be straight, meandering, anabranching (see ANABRANCHING AND ANASTOMOSING RIVER) or braided (see BRAIDED RIVER). To the extent that bedload transport dominates, and stabilizing, cohesive, floodplain sediments are lacking, gravel-bed channels tend to exhibit larger meander wavelengths and a propensity to wander or braid. Wandering is a type of anabranching that represents a transitional stage between meandering and braiding, with some sinuosity, low-level braiding and stable mid-channel islands. Wandering channels tend to have lower slopes and less abundant bedload than fully braided channels. Relative to sand-beds, gravel-bed channels require steeper slopes to generate full braiding.

Bed material size is a fundamental control of river form and function, and characteristic morphological and process attributes do justify the general distinction that is made by geomorphologists between gravel- and sand-bed rivers. The presence of a gravel–sand transition in many rivers and the widely reported deficiency of fluvial sediments in the range 1 to 4 mm – the so-called 'grain-size gap' – reinforce this binary categorization. However, all gravel-bed rivers contain sand, and many gravel-bed rivers transport larger volumes of sand than gravel. The sand is apparent to varying degrees as a patchy surface veneer (for example in pools) and in the subsurface matrix. This suggests that the two-fold, sand versus gravel, classification is rather simplistic and potentially limiting. It may obscure important attributes that are peculiar to channels containing particular mixtures of sand and gravel. Indeed, there is increasing evidence that understanding channel hydraulics, sediment transport and the formation of fluvial deposits in 'gravel-bed' rivers depends upon explicit recognition that bimodal gravel and sand mixtures often dominate the bed materials (e.g. Sambrook-Smith 1996).

References

Parker, G. (in press) Transport of gravel and sediment mixtures, in *Sedimentation Engineering*, American Society of Civil Engineers, Manual 54.
Sambrook-Smith, G.H. (1996) Bimodal fluvial bed sediments: origin, spatial extent and process, *Progress in Physical Geography* 20, 402–417.

Further reading

Simons, D.B. and Simons, R.K. (1987) Differences between gravel- and sand-bed rivers, in C.R. Thorne,

J.C. Bathurst and R.D. Hey (eds) *Sediment Transport in Gravel-bed Rivers*, 3–15, Chichester: Wiley.
Kleinhans, M.G. (2002) *Sorting Out Sand and Gravel: Sediment Transport and Deposition in Sand-gravel Bed Rivers*, Netherlands Geographical Studies 293, Utrecht: Royal Dutch Geographical Society.

SEE ALSO: armouring; bedload; downstream fining; roughness; sand-bed river

STEPHEN RICE

GRÈZE LITÉE

Stratified TALUS deposits displaying well-developed cm-thick beds and composed of small angular clasts (Guillien 1951). They have also been referred to as éboulis ordonnés and stratified screes, though some authors find slight differences between these terms (mostly related to the slope gradient and the mean clast size). The most diagnostic features of grèzes litées are (1) the internal structure of the deposit, organized in parallel beds of around 10 to 25 cm thick, and (2) the small size of the clasts as a result of very frequent freezing and thawing cycles on a gelivable (frost susceptible) rock substratum.

The sedimentary structure shows alternating matrix-rich (matrix-supported) and openwork (clast-supported) beds. In many cases openwork beds show fining upward textures. In longitudinal sections the base of the matrix-rich beds is affected by festoons, with increasing size downslope, and even with the development of lobate fronts (Bertran *et al.* 1992). Undulations are relatively frequent in frontal sections. The presence of blocks within the clast-supported beds defines another more heterogeneous type of talus deposit called groizes litées. In carbonate-rich deposits the presence of carbonate cemented crusts is relatively frequent, as a result of percolation and water circulation.

Most authors consider that grèzes litées are better developed in limestone areas, at the foot of large vertical or sub-vertical cliffs. This is the case of Charentes (France), where the best examples have been studied, and many other localities in the Alps and the Pyrenees. However, they have also been described in crystalline, volcanic and metamorphic rock areas (for instance, in the Chilean Andes, Vosges, the French Central Massif and the Atlas in Morocco). These deposits can be up to 40 m thick. In all cases the grèzes litées are most frequent in middle latitudes, with a periglacial climate. In these regions the annual number of freezing and thawing cycles is high (even more than 200 days per year), providing the best conditions to break down the rocks and to accumulate large volumes of debris. Under these conditions the cliffs erode backward rapidly and they are partially fossilized by the grèzes litées.

Grèzes litées have been described in a wide range of slope gradients (between 5 and 35°), though, in general, gentler than in ordinary talus with non-stratified screes, thus excluding an origin based only on gravity. Most active grèzes litées are located on sunny aspects or in snow-free hillslopes. Pleistocene deposits, related to former cold-climate phases, are located in almost any aspect, depending not only on the altitude but also on local topography and wind direction (García-Ruiz *et al.* 2001).

Several hypotheses have been used to explain the development of grèzes litées. Tricart and Cailleux (1967) stressed the importance of in-mass transport (especially solifluction) accompanied by pipkrake (needle-ice) activity. Bertran *et al.* (1992) and Francou (1988) confirm the decisive role of continual burial of stone-banked sheets. This implies the existence of large solifluction sheets in which pipkrakes cause a vertical sorting of the material, displacing the clasts towards the surface. The movement of the front of the stone-banked sheets produces the accumulation of continuous layers of clast-supported and matrix-supported beds. The slow mass movement is responsible for the occurrence of frontal and lateral festoons and undulations. The presence of debris flows also contributes to the characteristic alternating structure (Van Steijn *et al.* 1995), though the limits and continuity of the beds can be poorly developed. Slopewash processes are almost completely excluded as the main mechanism, since most of the rock fragments are oriented parallel to the slope gradient. Furthermore, the absence of rills, longitudinal sorting and cross-bedding suggests the inability of overland flow to redistribute the debris along the talus.

References

Bertran, P., Coutard, J.P., Francou, B., Ozouf, J.C. and Texier, J.P. (1992) Données nouvelles sur l'origine du litage des grèzes: implications paléoclimatiques, *Géographie Physique et Quaternaire* 46, 97–112.
Francou, B. (1988) Éboulis stratifiés dans les Hautes Andes Centrales du Pérou, *Zeitschrift für Geomorphologie* 32, 47–76.
García-Ruiz, J.M., Valero, B., González-Sampériz, P., Lorente, A., Martí-Bono, C., Beguería, S. and Edwards, L. (2001) Stratified scree in the Central Spanish Pyrenees: palaeoenvironmental implications, *Permafrost and Periglacial Processes* 12, 233–242.
Guillien, Y. (1951) Les grèzes litées de Charente, *Revue Géographique des Pyrénées et du Sud-Ouest* 22, 153–162.
Tricart, J. and Cailleux, A. (1967) *Le modelé des régions périglaciaires*, Paris: SEDES.
Van Steijn, H., Bertran, P., Francou, B., Hétu, B. and Texier, J.P. (1995) Models for the genetic and environmental interpretations of stratified slope deposits: review, *Permafrost and Periglacial Processes* 6, 125–146.

JOSÉ M. GARCÍA-RUIZ

GROUND WATER

Groundwater processes

Ground water is an important source of water for domestic use, irrigation and industrial uses and concern about the quantity and quality of ground water withdrawals is global in nature. It is a critical link in the hydrologic cycle, as it is a major source of water in rivers and lakes. Ground water is water under positive (greater than atmospheric) pressure in the saturated zone. The fluctuating water table marks the upper boundary of saturation in unconfined aquifers. Recharge can occur by infiltration of rainwater or snowmelt and by horizontal or vertical seepage from surface-water bodies. Ground water leaves the system by discharge into rivers, lakes or the ocean, by transpiration from deep-rooted plants, or by evaporation when the water table is close to the surface. Ground water is in continual motion, with velocities that are typically less than 1 m day^{-1}.

The most important geologic factors controlling the movement of ground water are lithology, stratigraphy and structure, and combinations of these conditions produce a great variety of groundwater flow patterns. The term aquifer is used to define a geologic unit that can store and transmit enough water to be hydrologically or economically significant. Layers of rock which are impermeable are termed aquicludes and semi-permeable rocks, which retard the flow, are termed aquitards. An unconfined aquifer is open to the atmosphere and its hydrostatic level is the water table. In a confined aquifer water is held between confining layers (aquitards or aquicludes) and is not vertically connected to the atmosphere.

In humid regions, ground water may be an important contributor to streamflow, with water entering the channel by effluent seepage to form the baseflow discharge. If ground water input is significant, the streams will be characterized by relatively low temporal flow variability. By contrast, in arid regions streamflow often percolates into permeable beds to contribute to the water table. Such streams are referred to as influent.

Ground water as a geomorphological agent

Ground water is a significant geomorphological agent in many environments, both arid and humid, hot and cold. It influences cave formation in karst terrains; water chemistry and surface morphology of playas or PANS; the erosion of rock faces and formation of alcoves and caves; cliff retreat and mass movement; and canyon growth by basal sapping processes. Ground water can impede wind erosion in arid areas where the water table lies close to the surface and can affect dune type. Over time, the role of ground water in geomorphic development is strongly affected by fluctuations in the height of the water table which result from climate change and human pumpage. Excessive groundwater withdrawal and falling water tables can lead to surface subsidence, vegetation death and dune mobilization.

The term KARST is given to limestone terrains that include such distinctive landforms as caves, springs, blind valleys and dolines. The dominant erosional process is dissolution and the region is typified by lack of surface water and the development of stream sinks or dolines. A unique pattern of drainage results from karst processes. Solution creates and enlarges voids, which then integrate to allow the transmission of large amounts of water underground, thereby promoting further solution. In karst areas underground drainage is developed at the expense of surface flow networks. Solution and weakening of silicic rocks to form karst-like topography has also been noted in arid environments, such as the Bungle Bungle of north-western Australia, but it is uncertain whether such landforms are wholly or partly inherited from more humid climatic periods.

Ground water plays a role in mass movement and channel formation by the process known as sapping. Concentrated seepage caused by ground water convergence is capable of slowly eroding materials at valley head or cliff bases, undermining overlying structures, and causing failure and headward retreat. The term spring sapping is often used when a point-source spring is involved, whereas seepage erosion may be employed where the groundwater discharge is less concentrated. Computer modelling suggests that scalloped escarpments develop where groundwater flow is diffuse, whereas elongation into channels or canyons results from higher and more concentrated seepage discharges, often associated with growth updip along fracture systems characterized by higher hydraulic conductivities. In rocks which are susceptible, chemical weathering renders the rocks even more permeable. Although many sapping networks, for instance those of coastal Italy and the Colorado Plateau, have developed in highly jointed bedrock, field research by Schumm *et al.* (1995) illustrates that similar networks can develop in highly permeable sands without significant structural controls.

Common morphological characteristics of valleys in which sapping plays a dominant role include amphitheatre-shaped headwalls, relatively constant valley width from source to outlet,

Plate 57 Long Canyon and Cow Canyon are tributaries to the Colorado River, developed in the Navajo Sandstone. The morphology of these valleys, with theatre-shaped heads and relatively constant valley width from source to outlet, is consistent with their formation by groundwater sapping

high and often steep valley sidewalls, a degree of structural control, short and stubby tributaries, and a longitudinal profile which is relatively straight (Plate 57). Similarly, simulation models (Howard 1995) of groundwater sapping produce canyons which are weakly branched, nearly constant in width and terminate in rounded headwalls.

Interest in groundwater outflow processes as an important factor in valley network development was stimulated by imagery of Mars which revealed features that were broadly similar in morphology to those on Earth (Laity and Malin 1985). It is now widely believed that many valleys on Mars were probably the result of erosion by groundwater sapping (Gulick 2001), although the actual mechanism is still subject to conjecture and debate. On Earth, an ever-increasing number of research papers illustrate that groundwater sapping is a global process, which can occur in a number of diverse lithologic and hydrologic settings. Valleys formed by sapping processes have been identified in Libya, Egypt, England, the Netherlands, the United States (Vermont; the Colorado plateau; Hawaii; Florida), New Zealand, Japan and Botswana. Valleys and escarpments which maintain the characteristic forms outlined above, but which lack modern seepage, may be relict from previously wetter climates (for instance, in Egypt). Other systems may include both active and relict components.

In addition to forming valleys by headward erosion, sapping at zones of groundwater discharge also contributes to the backwasting of scarps. Slopes are undermined and collapse owing to the removal of basal support by fluid flow which weakens rock at sites of concentrated seepage or diffuse discharge. These processes have received particular attention when considering scarp retreat in sandstone-shale sequences of the American south-west. Additionally, slopes of sandstone, granite, tuff or other massive rock form may be modified to include alveolar weathering or tafoni. The term 'dry sapping' has been applied to the formation of these features, for although the rock surfaces may be encrusted by salts, they do not appear to be damp. By contrast, larger alcoves formed by 'wet sapping' show wet surfaces on at least a seasonal basis (Howard and Selby 1994).

Boulders and inselberg landscapes in arid regions, when exposed by mantle stripping, are collectively referred to as etch forms. Such landforms may have developed over periods of 100 My or more and had an origin beneath deep mantles (tens or hundreds of metres in depth) which weathered as a response to ground water, under the control of geothermal heat. It has been proposed that the residual rock masses formed at the basal weathering front and were later exposed as the regolith was stripped.

Playa surfaces in deserts vary considerably owing to the range of unique hydrologic environments. Where ground water discharges seasonally or perennially onto the playa surfaces salt crystallization is characteristic and salt crusts of varying thickness form. The surface expression of groundwater discharge and salt crystallization includes extremely irregular micro-topography, polygonal forms and salt ridges. Solution phenomena such as pits and sinkholes may occur. Beneath the surface, the sediments are usually wet, soft and sticky. Spring mounds, elevated forms which may have a central pool, form where the water table is higher than the playa surface.

Landscape changes associated with groundwater overdraft

When more water is withdrawn from an aquifer by pumping than can be returned by natural recharge, the system is considered to be in overdraft. Such conditions have geomorphic impacts. Overdraft conditions have led to measurable SUBSIDENCE of the ground (one to ten metres) in such areas as Mexico City, Tokyo, Hanoi, the Central Valley of California, the Houston–Galveston area of Texas, and Las Vegas, Nevada. Surface geomorphic expression includes the development of fissure systems, such as those at Yucca dry lake, Nevada, where parallel fissures are as much as 2 km long and perhaps 500 m deep. In Bangkok, Thailand, subsidence averages 1.5–2.2 cm/yr, but has occurred at rates as high as 10 cm/yr, causing damage to buildings and infrastructure. As the city is almost at sea level, the most serious impact has been flooding at the end of the rainy season.

Ground water plays a significant role in aeolian and fluvial systems of deserts. In channel systems, ground water forced to the surface by faulting or bedrock may flow at the surface for short distances in zones marked by dense phreatophytic vegetation – plants whose root systems draw water directly from ground water, and which dominate the riparian habitat. Phreatophytes affect hydraulic roughness and depositional processes, and their loss owing to water table drawdown often precedes episodes of stream widening.

Dune systems also change in response to a decline in water table elevation. In a 'wet aeolian system' the water table lies at or close to the surface and various stabilizing agents, such as vegetation, deflation lags, or cements allow accumulation while the system remains active. Drawdown of the water table may lead to a change to a 'dry aeolian system', where neither the water table nor vegetation exerts any significant influence, and surface behaviour is largely controlled by aerodynamic configuration (Kocurek 1998). In the Mojave Desert of California, sand released from degrading nebkhas (vegetation-anchored dunes) has reaccumulated downwind in migrating sand streaks and barchan dunes. Problems to nearby settlement include dune encroachment and blowing dust episodes.

References

Gulick, V.C. (2001) Origin of the valley networks on Mars; a hydrological perspective, *Geomorphology* 37, 241–268.

Howard, A.D. (1995) Simulation modeling and statistical classification of escarpment planforms, *Geomorphology* 12, 187–214.

Howard, A.D. and Selby, M.J. (1994) Rock Slopes, in A.D. Abrahams and A.J. Parsons (eds) *Geomorphology of Desert Environments*, 123–172, London: Chapman and Hall.

Kocurek, G. (1998) Aeolian system response to external forcing factors – a sequence stratigraphic view of the Sahara Region, in A.S. Alsharhan, K.W. Glennie, G.L. Whittle and C.G. St C. Kendall (eds) *Quaternary Deserts and Climatic Change*, 327–337, Rotterdam: A.A. Balkema.

Laity, J.E. and Malin, M.C. (1985) Sapping processes and the development of theater-headed valley networks in the Colorado Plateau, *Geological Society of America Bulletin* 96, 203–217.

Schumm, S.A., Boyd, K.F., Wolff, C.G. and Spitz, W.J. (1995) A ground-water sapping landscape in the Florida Panhandle, *Geomorphology* 12, 281–297.

Further reading

Kochel, R.C. and Piper, J.F. (1986) Morphology of large valleys on Hawaii; evidence for ground water sapping and comparisons with Martian valleys, *Journal of Geophysical Research* 91, E175–192.

Laity, J.E. (in press) Ground water drawdown and destabilization of the aeolian environment in the Mojave Desert, California, *Physical Geography*.

Luo, W., Arvidson, R.E., Sultan, M., Becker, R., Crombie, M.K., Sturchio, N. and Zeinhom, E.A. (1997) Ground-water sapping processes, Western Desert, Egypt, *Geological Society of America Bulletin* 109, 43–62.

Péwé, Troy L. (1990) Land subsidence and earth-fissure formation caused by ground water withdrawal in Arizona, in C.G. Higgins and D.R. Coates (eds) *Ground Water Geomorphology; The Role of Subsurface Water in Earth-surface Processes and Landforms*, Geological Society of America Special Paper 252, 218–233.

Young, R.W. (1987) Sandstone landforms of the tropical East Kimberley region, northwestern Australia, *Journal of Geology* 95, 205–218.

SEE ALSO: canyon; etching, etchplain and etchplanation; karst; pan; sandstone geomorphology; spalling

JULIE E. LAITY

GROYNE

Groynes are shore-perpendicular structures that are emplaced to control sand movement along a beach by altering processes in the swash and surf zones and providing a physical barrier to sediment moved as littoral drift.

Groynes change patterns of wave-refraction, wave-breaking and surf-zone circulation, generate rip currents, trap sediments on the updrift beach, reduce sediment inputs to the downdrift beach and redirect sediment offshore. Geomorphic effects include creation of wider beaches with steeper foreshores on the updrift sides, narrower beaches with flatter foreshores on the downdrift sides, lobate deposition zones downdrift of the tips of the structures and pronounced breaks in shoreline orientation (Everts 1979). The locally wider updrift beaches can enhance aeolian transport, and the subaerial portions of the groynes can form effective traps for blown sand, increasing the potential for creation of dunes on their updrift side (Nersesian *et al.* 1992; Nordstrom 2000). However, shoreline recession rates may be greatly increased on the downdrift side of groynes (Everts 1979; Nersesian *et al.* 1992), leading to truncation of beaches and dunes and loss of habitat.

Shortening, lowering or notching of existing groynes or construction of permeable pile groynes or submerged groynes have been suggested to allow for some sediment to bypass the structures in order to reduce downdrift erosion rates. Permeable pile groynes can reduce the longshore current while eliminating the effect of the structurally induced rip current, creating a more linear shoreline than occurs with an impermeable groyne and creating an underwater terrace that can reduce the erosion potential of waves crossing it (Trampenau *et al.* 1996). Submerged groynes retain the original aesthetics of the landscape, and allow beach traffic to proceed unimpeded, but their effects have been poorly studied (Aminti *et al.* 2003). T-groynes, built with a short, shore-parallel seaward end, are favoured in some areas to reduce scour and redirect rip currents, thereby reducing unwanted sedimentation offshore, but they can leave the beach in the centre badly depleted (McDowell *et al.* 1993).

Groynes can be used to best advantage when they are located where (1) sediment transport diverges from a nodal region; (2) there is no source of sand, such as downdrift of a breakwater or jetty; (3) transport of sand downdrift is undesirable; (4) the longevity of BEACH NOURISHMENT must be increased; (5) an entire reach will be stabilized; and (6) currents are especially strong at inlets (Kraus *et al.* 1994). Groynes also have considerable recreational value for fishing because they create new habitat and provide access to deep water. Combined pier/groyne structures have been built to enhance this value.

New groynes or alterations to existing groynes are now often included in beach nourishment plans, but groynes have been banned or strongly discouraged in some management policies (Truitt *et al.* 1993; Kraus *et al.* 1994). Instances of removal of groynes have been reported (McDowell *et al.* 1993), but there is little documentation of the results on beach change. Alterations to groynes to allow for some bypass of sediment are more common than removal and are better documented (Rankin and Kraus 2003).

References

Aminti, P., Cammelli, C., Cappietti, L., Jackson, N.L., Nordstrom, K.F. and Pranzini, E. (2003) Evaluation of beach response to submerged groin construction at Marina di Ronchi, Italy using field data and a numerical simulation model, *Journal of Coastal Research*, Special Issue, in press.

Everts, C.H. (1979) Beach behaviour in the vicinity of groins – two New Jersey field examples, *Coastal Structures* 79, 853–867, New York: American Society of Civil Engineers.

Kraus, N.C., Hanson, H. and Blomgren, S.H. (1994) Modern functional design of groin systems, *Coastal Engineering: Proceedings of the Twenty-fourth Coastal Engineering Conference*, 1,327–1,342, New York: American Society of Civil Engineers.

McDowell, A.J., Carter, R.W.G. and Pollard, H.J. (1993) The impact of man on the shoreline environment of the Costa del Sol, southern Spain, in P.P. Wong (ed.) *Tourism vs Environment: The Case for Coastal Areas*, 189–209, Dordrecht: Kluwer Academic Publishers.

Nersesian, G.K., Kraus, N.C. and Carson, F.C. (1992) Functioning of groins at Westhampton Beach, Long Island, New York, *Coastal Engineering: Proceedings of the Twenty-third Coastal Engineering Conference*, 3,357–3,370, New York: American Society of Civil Engineers.

Nordstrom (2000) *Beaches and Dunes of Developed Coasts*, Cambridge: Cambridge University Press.

Rankin, K.L. and Kraus, N.C. (eds) (2003) Functioning and design of coastal groins: the interaction of groins and the beach: processes and planning, *Journal of Coastal Research*, Special Issue, in press.

Trampenau, T., Göricke, F. and Raudkivi, A.J. (1996) Permeable pile groins, *Coastal Engineering 1996: Proceedings of the Twenty-fifth International Conference*, 2,142–2,151, New York: American Society of Civil Engineers.

Truitt, C.L., Kraus, N.C. and Hayward, D. (1993) Beach fill performance at the Lido Beach, Florida groin, in D.K. Stauble and N.C. Kraus (eds) *Beach Nourishment: Engineering and Management Considerations*, 31–42, New York: American Society of Civil Engineers.

KARL F. NORDSTROM

GRUS

A product of *in situ* GRANULAR DISINTEGRATION of coarse-grained rocks characterized by its specific grain-size distribution, where the sand (0.1–2.0 mm) and gravel (> 2.0 mm) fraction predominate and may constitute up to 100 per cent of the total. The percentage of finer particles liberated by weathering is often negligible (Migoń and Thomas 2002). Thus, grus is not associated with any particular bedrock, although some rocks, e.g. mudstones, are unlikely to produce grus because of their grain-size composition. Granitic rocks, gneiss and migmatites are parent rocks that typically break down into grus.

The term is also used by sedimentologists to describe a product of accumulation of weathering-derived, poorly sorted, angular quartz and feldspar grains that have been subjected to very limited transport, usually towards the base of an outcrop. Such a sedimentary veneer of grus is particularly widespread in arid and semi-arid areas, where slope wash redistributes products of rock disintegration across PEDIMENTS.

Grus as the product of current superficial weathering of rock outcrops should not be confused with grus weathering mantles, which may be many metres thick and can be found in geological records. Grus saprolites may be defined as *in situ* weathering profiles, consisting almost entirely, or predominantly, of grus throughout, that grades into unweathered parent rock. Grus may occur at the base of a deep weathering profile and would represent a transitional stage in alteration of solid rock into a clayey weathering mantle, although there is evidence that many tropical deep weathering profiles do not have a basal zone of grusification and the transition zone is very thin. 'Arenaceous mantles' and 'sandy saprolites' are usually used as synonyms of grus mantles.

Grus saprolites are diversified in terms of their internal structure, depth and lithology. Many of them are homogeneous throughout, yet some contain frequent core stones, zones of more advanced breakdown along fractures or show sharp lateral or vertical contacts between weathered and unweathered parent rock. Core stones within grus profiles may be as large as 3–4 m across and be either closely spaced, separated by weathered fractures, or in isolation in an otherwise strongly disintegrated rock mass. Various depths of grus saprolites are reported and profiles more than 10 m deep are not uncommon. Mineralogical changes associated with grusification are usually slight and the content of secondary clay minerals in grus profiles is often insignificant (<2 per cent). Among clays, interstratified minerals, kaolinite, halloisyte and vermiculite are the most common. The occurrence of gibbsite is reported, but its percentage is usually low and is likely to represent a transitional stage in the formation of kaolinite.

The origin of deep grus saprolites is still unclear and several mechanisms have been suggested to be responsible for opening of microfractures within and between the grains in the near-surface zones (Pye 1985; Irfan 1996). Microfracturing results from de-stressing of quartz and feldspars during weathering and may be enhanced by expansion of biotite after its HYDRATION. Development of intergranular porosity in response to partial solution along grain boundaries and transgranular microcracks may be an important contributing agent. However, advanced chemical processes play rather a subordinate, if any, role as indicated by minor amounts of secondary clay, preservation of easily weatherable minerals such as biotite and plagioclase, and limited degree of corrosion of quartz and feldspar grains.

Grus mantles are particularly widespread in areas of temperate climate, but they in fact occur in a variety of climatic zones and may be found in every climatic regime, both humid and semi-arid (Migoń and Thomas 2002). In low latitudes they occur alongside products of more advanced alteration, such as ferrallitic saprolites, as for instance in south-east Brazil. This distribution contradicts the claim, often made in the past, that production of grus is primarily controlled by climatic conditions, and that it is specific for a humid temperate climate. Generalization is further inhibited by the possibility that many grus mantles are not the result of weathering under contemporary climatic conditions, but are inherited from a geological past and different climatic regimes, and by the fact that many grus profiles are evidently truncated.

From a geomorphological point of view, grus mantles occur in three major settings. First, they are common within elevated plateaux and uplands, beneath gentle upper slopes and along valley sides. Second, they occur in hilly and inselberg landscapes, but hills may either be weathered throughout into grus or have only their lower slopes underlain by a grus mantle. Third, they are associated with highly dissected mountain areas. In some subtropical mountains, watershed ridges, spurs and isolated hills are very often deeply weathered and only the cores of unweathered bedrock protrude as massive domes from the widespread saprolitic mantle (Thomas 1994).

The common association of thick grus mantles with areas of moderate to high relief, often with a recent history of uplift, across the world's morphoclimatic belts, implies that dissected terrains of moderate relief are particularly suitable for thick grus to develop. This is because of free drainage, strong hydraulic gradient, tensional stress and rock dilatation. On the other hand, surface instability prevents grus profiles from attaining geochemical and mineralogical maturity. It has therefore been proposed that the deep grus phenomenon is a response of weathering systems to rapid relief differentiation, whether by tectonics, erosion, or both, and associated enhancement of groundwater circulation, although it is not exclusive to such settings (Migoń and Thomas 2002).

References

Irfan, Y.T. (1996) Mineralogy, fabric properties and classification of weathered granites in Hong Kong, *Quarterly Journal of Engineering Geology* 29, 5–35.

Migoń, P. and Thomas, M.F. (2002) Grus weathering mantles – problems of interpretation', *Catena* 49, 5–24.

Pye, K. (1985) Granular disintegration of gneiss and migmatites, *Catena* 12, 191–199.

Thomas, M.F. (1994) *Geomorphology in the Tropics*, Chichester: Wiley.

Further reading

Dixon, J.C. and Young, R.W. (1981) Character and origin of deep arenaceous weathering mantles on the Bega batholith, Southeastern Australia, *Catena* 8, 97–109.

Lidmar-Bergström, K., Olsson, S. and Olvmo, M. (1997) Palaeosurfaces and associated saprolites in southern Sweden, in M. Widdowson (ed.) *Palaeosurfaces: Recognition, Reconstruction and Palaeoenvironmental Interpretation*, Geological Society Special Publication 120, 95–124.

SEE ALSO: granite geomorphology; weathering

PIOTR MIGOŃ

GULLY

'Gully' can refer correctly, if uncommonly, to clefts down cliffs and to several sorts of seafloor channels. Those uses stem logically from the root sense of a narrow passageway, following derivation from the Latin *gula*, meaning throat, the French *goulet*, meaning a narrow entry or passage (including into a harbour or bay), and the Middle English *golet*, or gullet.

Minor uses aside, however, *gully* predominantly denotes a small and narrow but relatively deeply incised stream course, difficult to cross or to ascend, for which words like valley and gorge are too grandiose. It ideally connotes a young cut, with steep sides and a steep headwall, that has been carved out of unconsolidated regolith, typically by ephemeral flow from rainstorms or meltwater. However, these are not required attributes and exceptions abound.

Gullies are very variable in terms of processes of initiation and growth, as well as conditions of substrate, vegetation and climate, so they vary greatly in appearance and can show distinct regional differences. Thus the literature contains diverse usages and many local synonyms, including DONGA, vocaroca, ramp and lavaka. Among the variations, gullies may be slit-like to lobate (expanded at the head end), and continuous or discontinuous (depending on whether or not the gully has become connected at grade to the main drainage system). They can grow downward from mid-hillslope positions or from the hill-toe up, or they can develop along valley floors. Most gullies have a relatively simple, single thalweg, but gully heads can split during headward retreat, thereby creating a dendritic shape, and once in a while two branches rejoin head-to-head, creating a ring valley around a central pedestal or hillock.

Gully has distinct but poorly defined connotations of size, and even vaguer implications of cross-sectional shape. A gully is bigger than a RILL, which is a small entrenched rivulet, small enough to be crossed by a wheeled vehicle or to be eliminated by ploughing. An ARROYO (or wadi or barranca) represents an entrenched stream, not necessarily very deep, that has a somewhat wider floor than a gully – one might hope to drive up an arroyo, but probably not up a gully. A gully has a greater width to depth ratio than a slot canyon, and is neither as deep nor as wide as a box canyon or a gorge (see GORGE AND RAVINE), all of which would also likely have rock walls, unlike typical gullies. A gully is ideally narrower and shallower than a ravine, although no size limits have been specified. Gullies and gorges can share equally steep and enclosing walls, but ravines can be more V-shaped in cross section. Floors of ravines are ideally less enclosed than floors of gullies, but need not offer easier access. Ravines can be cut in regolith or rock. Overall, gully and ravine overlap considerably (the French *ravine* explicitly includes both gullies and larger valleys). In popular usage, gullies, ravines and gorges are perhaps best separated by their implications of lethality: a fall into a gorge could easily be fatal, whereas only the terminally unlucky would die by falling into a gully, and falls into ravines are unpredictable. Gullies might therefore be considered to range approximately from 5 or 10 m long, 1 or 1.5 m wide, and about as deep, arguably up to the order of several hundred metres long, many tens of metres wide at the original ground level, and perhaps twenty or thirty metres deep. Use of 'gully' at the larger end of that spectrum seems most supportable when there is a gradation in size from similar but smaller gullies nearby.

Probably the most dramatic and common causes of gullies involve human misuse of the land, where sites are made vulnerable to erosion by deforestation or by a more general devegetation, via logging, burning, overgrazing, or establishment of fields (especially when on hillsides and when unterraced or ploughed down the slope rather than across it). Additional proximal causes related to humans include runoff along paths or tracks that run straight downhill, regrading of hillsides and diversion or concentration of runoff from roads or building sites uphill. Gullying can also be initiated when soil compaction tips the balance from infiltration to runoff, or following mass movement after a hillside is undercut or overloaded.

However, there are also many natural causes for gullying. Exceptionally intense or prolonged storms are a very common culprit in erosion, but critical increases in rainfall can also come about from such climate changes as increased annual rainfall, or increased storminess with no net increase in rainfall. Critical levels of devegetation can be reached by aridification or by natural fires.

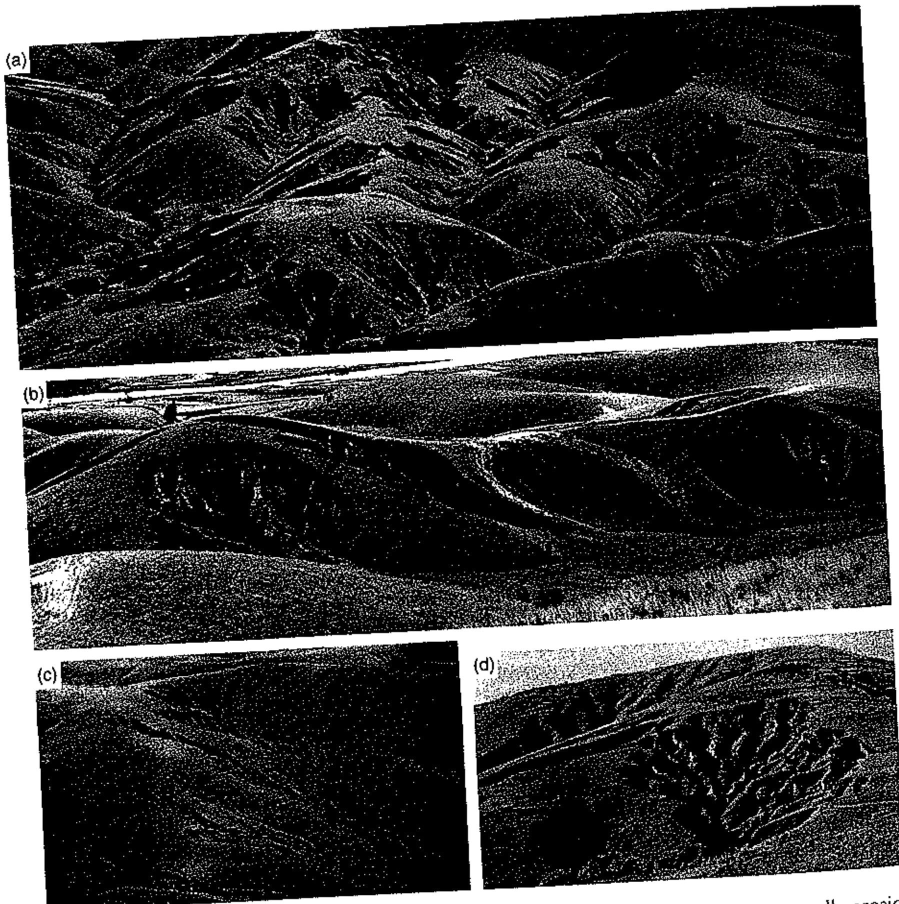


Plate 58 Examples of gully erosion in hills of weak saprolite in Madagascar. (a) Intense gully erosion. Concave-up runoff profiles are replacing smoothly convex hill profiles that formed by infiltration and chemical weathering. (b) Gullies at various stages of evolution. The biggest gully has expanded headward up dip, through the ridge crest. Ridge-crest cattle trails attest to endemic overgrazing in Malagasy hills. Many of these gullies receive no runoff from upslope. (c) These long and narrow gullies apparently represent entrenchment of a pre-existing dendritic drainage system. (d) Erosion dominated by runoff generated within the gully

Gully erosion may also be initiated by natural slope collapse, varying in style from deep slumps to soil slips, which can be triggered by rainstorms, earthquakes or undercutting by springs or rivers. Another cause is the collapse of pipes (see PIPE AND PIPING), which are natural subsurface passageways through soils, enlarged by rapid drainage along burrows, root voids, interconnected soil pores and the like. Critical increases in runoff relative to infiltration may happen naturally due to the plugging of a soil with fines or (especially in laterites) hardening due to exposure and drying following devegetation. Gullying can also be caused by incision or headward extension of first-order streams, as a result of uplift of headwater regions, subsidence downstream, a fall in base level, an increase in runoff and discharge, or breaching of a dam or a sill.

Note that the causes that make a site vulnerable to gullying can be very different from the specific initiators of erosion, such as when fires remove vegetation, but erosion starts only when sufficiently intense or prolonged rains attack the site. Thus, gullies may easily have different proximal and ultimate causes, and causes may operate at the gully, or upstream or downstream. Thus also, many gullies show threshold-related behaviour rather than linear responses to processes. Overall, gullying tends to be diagnostic of recently disrupted or non-equilibrium landscapes, whether perturbed by natural or cultural agents.

Once erosion is initiated, it can continue by a variety of processes, and the processes may change during growth. Rain attack and slope wash on the walls can be surprisingly effective, as can spalling or collapse of the walls from multiple wetting/drying cycles. Stream incision along the gully floor and concomitant undercutting of sidewalls are major growth processes. Another very effective process is headward retreat of a waterfall at the head end, which is notably associated with gullies that formed due to a lowering of base level. However, some gullies form entirely from rain that falls within the gully itself and have little or no exterior catchment area and no streams or rivulets flowing into them. In some situations, such as in thick saprolite in Madagascar, once the gullies have cut deeply enough to intersect wet zones near bedrock, groundwater seepage out of the bases of the headwall and adjacent sidewalls keeps the bases of those parts of the walls moist and vulnerable to further erosion and spalling, hence causing a dramatic increase in growth around the head end (Plate 58). This in turn creates very distinctive lobate or teardrop shapes, with broad arcuate heads and tiny exits. In these instances, the ultimate size of a gully is determined by the limits of the supply of subsurface water rather than the limits of the watershed on the surface. Gullies that grow by seepage and sapping can in some situations cut far back into high ground, effectively becoming amphitheatre-headed valleys. On occasion, they can even grow backward through ridge crests, for example when layering in the regolith dips through a ridge and delivers GROUND WATER from one side of a hill to the other.

In most cases, the shape and position of a gully reflect its causes and growth processes. For example, collapse of pipes, downcutting along stream beds, and headward retreat by springs or waterfalls create very long gullies, whereas growth by seepage and sapping can lead to a lobate shape with a broad and arcuate headscarp. Gullies that have grown along tracks and paths may lie along the crests of hill spurs, if those provided the easiest routes up the hill.

Most valley-side gullies represent new or future extensions of drainage networks into hillsides. Thus gullies commonly cut deeply into high ground, but they typically start within it and rarely pass through it at a low level, unlike most gorges. However, arroyo-like valley-floor gullies (which could form in the floor of a gorge) typically represent renewed incision of pre-existing drainages. Because they form most easily in unconsolidated material, gullies are common not only in colluvium on hillsides and alluviated valley floors but also in loess, till, outwash, loose fine volcaniclastics, laterite, saprolite and anthropogenic fill. The low resistance to erosion of these materials means that long-lived gullies are most typically associated with infrequent flow, which gives gullies a common but non-causal and non-exclusive association with relatively dry climates. (In contrast, ravines are more likely to have resistant walls and permanent flow.) It also means that gullies can grow extraordinarily rapidly. Thick regolith is generally a prerequisite for impressively deep gullies.

Left to themselves, gullies will eventually stop expanding, albeit possibly only when they have consumed their entire upslope watershed or have run out of erosible material. At that time, it will either have become a box canyon or ravine, or, if still cut into regolith, the upper walls will crumble back and the floor and lower walls will fill in and be buried, and the whole will become overgrown. Such a gully will become less angular and more rounded, with smoothly concave longitudinal and transverse profiles. Most will ultimately evolve into minor hillside hollows or reentrants.

Possibilities for remediating a gully include diverting drainage away from its head end; revegetating its walls and floors and/or the surrounding hillside; contour ploughing the surrounding hillside to promote infiltration rather than runoff; establishing small sediment-trapping dams along the floor, filling it in, and regrading the entire hillside. Nevertheless, the remedy should be matched to the specific causes and growth processes dominating each gully: disturbance of the surface (e.g. contour ploughing) will not be helpful if removal of vegetation is the critical initiator of erosion, and diversion of surface drainage may merely create a bigger problem somewhere else. If the gully is growing by sapping at the base of the walls, then work on the upper walls and the surrounding hillside may be at best a waste of energy, and effort should instead be concentrated on burying the wet zone, for example by promoting deposition along the gully floor. Overall, gullies are easier to prevent than to cure.

Further reading

Harvey, M.D., Watson, C.C. and Schumm, S.A. (1985) Gully erosion, *US Department of the Interior, Bureau of Land Management, Technical Note* 366, 1–181.
Higgins, C.G. (1990) Gully development, in C.G. Higgins and D.R. Coates (eds) *Groundwater Geomorphology*, Geological Society of America Special Paper 252, 18–58.
Ireland, H.A., Sharpe, C.F.S. and Eargle, D.H. (1939) Principles of gully erosion in the Piedmont of South Carolina, *US Department of Agriculture Technical Bulletin* 633.
NWSCA (National Water and Soil Conservation Authority) (1985) Soil erosion in New Zealand, *Soil and Water* 21(4), supplement.
Wells, N.A, and Andriamihaja, B. (1993) The initiation and growth of gullies in Madagascar: are humans to blame?, *Geomorphology*, 8, 1–46.

NEIL A. WELLS

GUYOT

Although occasionally used to refer to any sizeable underwater ocean-floor edifice, the term 'guyot' should be confined to those that are flat-topped and were once above the ocean surface. This is to distinguish them from seamounts which are underwater volcanoes that have never been above the ocean surface.

The flat tops of guyots were thought for many years to be erosional – an expression of wave truncation of their summits during an island's slow submergence. The first to be studied in detail were in the Hawaiian chain, and submergence and summit truncation were thought to be natural and unavoidable consequences for an island moving along the chain (Hess 1946).

As ideas of Earth-surface mobility became fashionable in the 1960s, so it was realized that the distribution of guyots about mid-ocean ridges (seafloor spreading centres) was significant. Most such guyots had evidently originated at the mid-ocean ridge and then, following a period as a subaerial volcanic island (or atoll), they were submerged as they moved down the ridge's steep flanks and became guyots.

Another important step in the understanding of the significance of guyots came when they were found to be mixed in with atolls in various Pacific island groups like the Marshall Islands, Austral, Tuamotu and several in Kiribati (and some in the Indian and Atlantic Oceans). It has become clear that these guyots were once ATOLLS. That they are no longer is due to various reasons, including the morphology of the ocean floor in these regions (particularly the presence of intraplate swells) and oceanographic factors (principally temperature) which inhibit coral growth (Menard 1984).

Guyot morphology and location

The existence of low-relief surfaces on the summits of guyots is clear evidence for most authors of wave truncation, specifically shoreline erosion (at typical rates of 1 km/Ma) coincident with island subsidence (e.g. Vogt and Smoot 1984). Coral reefs on some guyots demonstrate that they are drowned atolls (see below). Phosphorites on certain Pacific guyots (at depths of 550–1,100 m) also derive from subaerial avian phosphorites (Cullen and Burnett 1986) demonstrating that these islands were once above the ocean surface.

Following observations and insights of Charles Darwin, it was proposed in 1982 that an oceanographic threshold existed in the Hawaiian chain which explained why atolls became converted to guyots at around 29 °N. It was proposed that at the 'Darwin Point', the gross carbonate production by corals was no longer sufficient for atoll reefs to regrow during periods of sea-level rise and so the atoll which had once existed became drowned (Grigg 1982). More recently it has been argued that other factors such as climate and sea-level history, palaeolatitude, seawater temperature and light all contribute to the Darwin Point which has shifted in the Hawaii region between 24 and 30°N within the last 34 Ma (Flood 2001).

Although guyots are commonly located at the older ends of hotspot island chains where these cross the Darwin Point, other guyots are located in equally instructive locations. For example, the morphology of guyots which have been pulled down into the Tonga–Kermadec Trench (southwest Pacific) has given us insights into the nature of tectonic processes across oceanic plate convergent boundaries (Coulbourn *et al.* 1989).

References

Coulbourn, W.T., Hill, P.J. and Bergerson, D.D. (1989) Machias Seamount, Western Samoa: sediment remobilisation, tectonic dismemberment and subduction of a guyot, *Geo-Marine Letters* 9, 119–125.
Cullen, D.J. and Burnett, W.C. (1986) Phosphorite associations on seamounts in the tropical southwest Pacific Ocean, *Marine Geology* 71, 215–236.
Flood, P.G. (2001) The 'Darwin Point' of Pacific Ocean atolls and guyots: a reappraisal, *Palaeogeography, Palaeoclimatology, Palaeoecology* 175, 147–152.
Grigg, R.W. (1982) Darwin Point: a threshold for atoll formation, *Coral Reefs* 1, 29–34.
Hess, H.H. (1946) Drowned ancient islands of the Pacific Basin, *American Journal of Science* 244, 772–791.
Menard, H.W. (1984) Origin of guyots: the *Beagle* to *Seabeam*, *Journal of Geophysical Research* 89(B13), 11,117–11,123.
Vogt, P.R. and Smoot, N.C. (1984) The Geisha Guyots: multibeam bathymetry and morphometric interpretation, *Journal of Geophysical Research* 89(B13), 11,085–11,107.

PATRICK D. NUNN

GYPCRETE

The accumulation of gypsum ($CaSO_4 \cdot 2H_2O$) within a soil or sediment profile leads to the formation of gypsic horizons, which are called gypcrete. Gypcrete is a member of the dryland DURICRUST family, which also includes CALCRETE and SILCRETE. As gypcrete is more soluble than other duricrusts it seldom produces MESA landscapes with a gypsum CAPROCK. The process of gypcrete formation and its global distribution differs from that of other duricrusts. Nevertheless, calcrete and gypcrete may occur in the same profile and can be associated with salts such as halite. Gypcrete may also be host to one of the most commonly recognized forms of gypsum which is the desert rose.

Gypsum, the building material in the formation of gypcrete, is a very common mineral. It can be found throughout the world and forms under present-day evaporitic conditions in inland salt lakes (PANS), coastal salt flats (SABKHAS), springs, CAVES, organic rich submarine sediments and in dryland soils. It is most commonly associated with massive sedimentary bedrock sequences, in particular those of evaporitic lake or sea basins. Gypsum formation generally requires evaporation of water and due to its solubility forms preferentially in areas of very low rainfall. It may also produce GYPSUM KARST if massive gypsum horizons are subjected to significant precipitation. Surficial gypsum does occur in all arid regions including the polar deserts, but massive gypcrete appears to be restricted to the drier subtropical desert regions of North Africa, the Middle East, southwestern Africa, southwestern America, South America, Central Asia and Western Australia.

Significant primary sources of gypsum formation are pans and sabkhas, which may also host gypcrete. These environments often feature shallow groundwater tables that are subject to substantial evaporation rates, which lead to the formation of evaporites above the water table in close proximity or at the surface. An upward migration of water and formation of gypsum is described as *per ascensum* gypcrete formation. When shallow ground water evaporates into a sandy substrate around a pan or sabkha margin, desert rose gypcrete forms. These crystals can be up to 30 cm in size and can be joined to form a single massive horizon.

Pans may also be subject to pronounced surface salinity gradients between the point of freshwater input and the point of saline brine formation at the centre. Such gradients are often accompanied by distinct evaporitic zones, which in a circular pan are arranged in concentric belts. Under such conditions, sulphate formation often follows the formation of carbonates and precedes the precipitation of chlorides and other salts. Pan surface gypsum may form hard gypsum crusts, which can develop a thrust polygon pattern.

Pan or sabkha environments may also accumulate unconsolidated fine gypsum crystals and powdery gypsum soils on their surface. Such freshly precipitated gypsum may not always form a hardened surface crust, but may be subject to aeolian dispersal (see AEOLIAN PROCESSES) Gypsum deflation may lead to gypsiferous lunette dunes, in particular at pan margins and will lead to the accumulation of gypsum-rich dust in the downwind environment. Aeolian dispersal of gypsum from pans is common and may be relatively rapid as indicated by significant burial of Roman artefacts in Tunisia (Drake 1997). It may also produce regional-scale gypcrete as demonstrated in the Namib Desert region (Eckardt *et al.* 2001).

Pedogenic accumulations form during sporadic rain, which dissolves surface dust and reprecipitates gypsum in stable soils or sediment profiles. Regolith cover in particular traps gypsum dust and gradually incorporates gypsum into the stable subsurface soil or sediments below the STONE PAVEMENT. It has been suggested that pavement surfaces are displaced upward during the process of gypsum dust entrapment (McFadden *et al.* 1987). The external and primary aeolian input of gypsum into a soil results in the relative downward migration of gypsum into the profile and is described as the *per descensum* mode of gypcrete formation. This process generally takes place in the absence of ground water. The resulting gypcrete can be massive in character, and may partly consolidate the surface regolith of a stone pavement. Stone pavement surfaces that cover significant gypcrete accumulations sometimes develop polygonal surface patterns.

The various crusts outlined above may differ considerably in terms of thickness, strength, composition and purity. The structure of gypcrete may range from powdery, nodular to massive horizons that vary in thickness from a few centimetres to many metres. Gypsum crystals may vary in size from microcrystalline (powdery) to massive (desert rose) and may include alabasterine morphologies as well as transparent lenticular clasts. A single crust may undergo multiple stages of formation with reworking, removal, production and storage of crust occurring simultaneously. This can produce a spatially complex and varied morphology. As a result no typical gypcrete profile exists. Gypsum crusts can however be defined as 'accumulations at or within 10 m of the land surface from 0.10 m to 5.0 m thick containing more than 15 per cent by weight of gypsum and at least 5.0 per cent by weight more gypsum than the underlying bedrock' (Watson 1985).

The formation of gypcrete is not only determined by climate but also by the provision of the elements, which produce gypsum. In particular sulphate is not as common as the elements required to form silcrete or calcrete. Sulphur isotopes have demonstrated that the formation of gypsum is dependent on the supply of dissolved sulphate or pre-existing sulphate accumulations such as bedrock gypsum. Gypcrete formation is thereby directly linked to the regional sulphur cycle (Eckardt 2001). Dissolved sulphate in surface or ground water leading to the formation of gypsum may be derived from the dissolution of sulphates or sulphides in the bedrock, the marine atmospheric contribution of sea spray, marine dimethyl sulphide (CH_3SCH_3) also known as DMS (Eckardt and Spiro 1999), gaseous hydrogen sulphide (H_2S) or the dissolution of aeolian sulphate from terrestrial gypsum dust sources.

We still have little information on exact rates of gypcrete formation and the response of gypcrete to climatic change (Plate 59). Attempts have been made to examine the micropetrography of gypcrete and to infer palaeoenvironmental conditions from such observations (Watson 1988). Dating of gypsum and gypcrete is possible using U-series dating and thermal luminescence techniques.

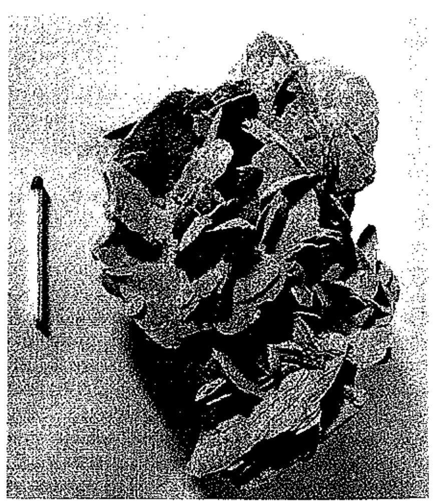


Plate 59 Example of a desert rose, the most commonly recognized form of gypsum. It forms with the evaporation of shallow ground water into a sandy substrate around a pan or sabkha margin

We have also been able to map gypcrete using remote sensing. Due to the distinct spectral response of gypsum in the mid-infrared region of the spectrum (2.08–2.35 μm) it can be separated from many other dryland features (White and Drake 1993).

Some powdery gypcrete accumulations may attain purities which make them attractive to mining while other deposits, in particular those fed by ground-water in Namibia and Australia, are known to be associated with high concentrations of uranium (Carlisle1978). In some areas pure gypcrete is also being used as a road paving material.

Gypcrete formation needs to be examined in the context of regional dryland processes, which include a combination of fluvial and aeolian processes. An understanding of pan or sabkha chemistry and hydrology is particularly important, as these are significant aeolian point sources of gypsum dispersal.

References

Carlisle, D. (1978) *The distribution of calcretes and gypsum in SW USA and their uranium favorability based on a study of deposits in Western Australia and South West Africa*, US Dept of Energy Subcontract, Open File Report, 76–022–E.

Drake, N.A. (1997) Recent aeolian origin of surficial gypsum crusts in southern Tunisia: geomorphological, archaeological and remote sensing evidence, *Earth Surface Processes and Landforms* 22, 641–656.

Eckardt, F.D. (2001) Sulphur isotopic applications, example: origin of sulphates, *Progress in Physical Geography*, 24(4), 512–519.

Eckardt, F.D. and Spiro, B. (1999) The origin of sulphur in gypsum and dissolved sulphate in the Central Namib Desert, Namibia, *Sedimentary Geology* 123 (3–4), 255–273.

Eckardt, F.D., Drake, N.A, Goudie, A.S. White, K. and Viles, H. (2001) The role of playas in the formation of pedogenic gypsum crusts of the Central Namib Desert, *Earth Surface Processes and Landforms* 26, 1,177–1,193.

McFadden, L.D., Wells, S.G. and Jercinovich, M.J. (1987) Influences of eolian and pedogenic processes on the origin and evolution of desert pavements, *Geology* 5(15), 504–508.

Watson, A. (1985) Structure, chemistry and origins of gypsum crusts in southern Tunisia and the central Namib Desert, *Sedimentology* 32, 855–875.

——(1988) Desert gypsum crusts as palaeoenvironmental indicators: a micropetrographic study of crusts from southern Tunisia and the central Namib Desert, *Journal of Arid Environments* 15(1), 19–42.

White, K.H and Drake, N.A. (1993) Mapping the distribution and abundance of gypsum in south-central Tunisia from Landsat Thematic Mapper data, *Zeitschrift für Geomorphologie* 30(3), 309–325.

Further reading

Cooke, R.U., Warren, A. and Goudie, A.S. (1993) *Desert Geomorphology*, London: UCL Press.

Watson, A. and Nash, D. (1997) Desert crusts and varnishes, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 69–107, Chichester: Wiley.

SEE ALSO: duricrust; gypsum karst; pan; sabkha

FRANK ECKARDT

GYPSUM KARST

Karst associated with gypsum and anhydrite rocks is generally referred to as 'Gypsum Karst' and has received little appreciation by geomorphologists if compared to the normal (limestone) KARST. However, gypsum karst is widely spread in the world where the global gypsum-anhydrite outcrop exceeds 7 million km^2, the largest areas being in the northern hemisphere, particularly in the United States, Russia and the Mediterranean basin.

Due to the high solubility of calcium sulphate, the gypsum karst life cycle is commonly far shorter than that of carbonate karst. The average experimental values for gypsum degradation within the Mediterranean area were 0.91 mm/1,000 mm of rain (Cucchi *et al.* 1998) and therefore no outcrop of such rock may survive more than a few hundred thousand years if exposed to the meteorological agents. The actual evolution depends greatly upon the geological history of the particular region so that intra-Messinian and even older gypsum karst may have been preserved until the present.

Exposed karst forms

Medium- to large-sized gypsum karst landforms (DOLINES, BLIND VALLEYS, polje-like depressions) are very similar in genesis and morphology to those found on carbonate rocks, while meso-, micro- and nano-forms may sometimes be peculiar to a gypsum environment. The differences in meso-, micro- and nano-forms are normally the direct consequence of the fact that the size of the crystals in different gypsum outcrops may range from over a metre to a fraction of a mm, while in the carbonate rocks the crystal size is normally around a mm.

The most peculiar gypsum karst meso-form are 'tumulos', while 'weathering crusts' are amongst the micro-forms. All of them develop in gypsum formations characterized by a crystal size of 1–10 cm, which is normal for the Messinian gypsum in the Mediterranean area. Their evolution is produced by the increase of volume of the superficial gypsum stratum induced by the dis-aggregation of the rock texture, as a consequence of the anisotropic behaviour of the gypsum crystals with respect to temperature changes (Calaforra 1998).

Finally, whereas in carbonate rocks some of the micro- and most of the nano-forms are the result of biological activity, the outcrop of very large gypsum crystals (up to 1 m or more in length) together with their high solubility allows for the evolution of nano-forms, the morphology of which is simply controlled by the structure of the crystal lattice (Forti 1996).

Deep karst forms

In gypsum karst the single active speleogenetic mechanism is simple dissolution. Therefore, here deep forms are not so varied as in carbonate karst, where plenty of different speleogenetic mechanisms are active. Moreover, most of the dissolution–erosion forms (pits, canyons, domes, scallops, large collapse chambers) are quite similar to those present in carbonate ones.

Gypsum CAVES are generally very simple linear or crudely dendritic caves that directly connect sink points and resurgence. They are commonly referred to as 'through caves' and consist of a principal drainage tube running along the water table with few and short, often subvertical, effluents; through caves are common in almost every entrenched and denuded gypsum karst area.

The deepest gypsum caves currently known rarely exceed 200 m in depth, being far shallower than those in carbonate karst: the reason is that always in mountainous regions, where the potential drained depth is greatest, gypsum formations are fragmented and do not favour the development of such vertically extensive sequences as do carbonates.

For the same reason the length of a gypsum cave rarely exceeds 2–5 km even if, in peculiar hydrogeological conditions (basal and/or lateral injection and dispersed inputs), complex dendritic 2- or 3-dimensional (multistorey) maze caves may develop up to several tens of kilometres. Podolia (Ukraine) is the 'type' region in which such caves have been explored and studied.

Chemical deposits

Chemical deposits (Hill and Forti 1997) are rather uncommon if compared with those present in carbonate caves: this depends mainly on the scarce chemical reactivity of gypsum. Normally they consist of calcite and gypsum and the local prevalence of one or the other mineral depends on climate.

Calcite SPELEOTHEMS show no morphologic peculiarities to distinguish them from similar deposits in limestone caves. Although, in most cases, their depositional mechanism is unlike that which dominates in a limestone environment (supersaturation due to CO_2 loss) being the product of the incongruent dissolution of gypsum by water with a high initial carbon dioxide content. Incongruent dissolution also explains the existence of unique forms like 'calcite blades' and 'half calcite bubbles'.

Gypsum speleothems have a more ubiquitous distribution. They present obvious morphological differences compared to calcite ones, due to their distinct genetic mechanism, which involves supersaturation due to evaporation. This genetic mechanism is also responsible for several unique forms such as 'gypsum balls', 'gypsum hollow stalagmites' and 'gypsum powder'.

Climatic influence on the chemical deposits

In gypsum caves climatic factors have a strong influence upon calcite and/or gypsum deposits. The completely different depositional mechanisms (incongruent dissolution for calcite and evaporation for gypsum) are influenced in very different ways by climatic variables: therefore climate strictly controls (far more than in carbonate karst) what chemical deposit can develop in a given gypsum cave. This close relationship with climate gives deposits preserved in the gypsum environment a potentially very great importance on the basis of their application to palaeoclimatology (Figure 79) and as indicators for present-day climatic changes.

References

Calaforra, Chordi, J.M. (1998) *Karstologia de yesos*, Universidad de Almeria, Spain.

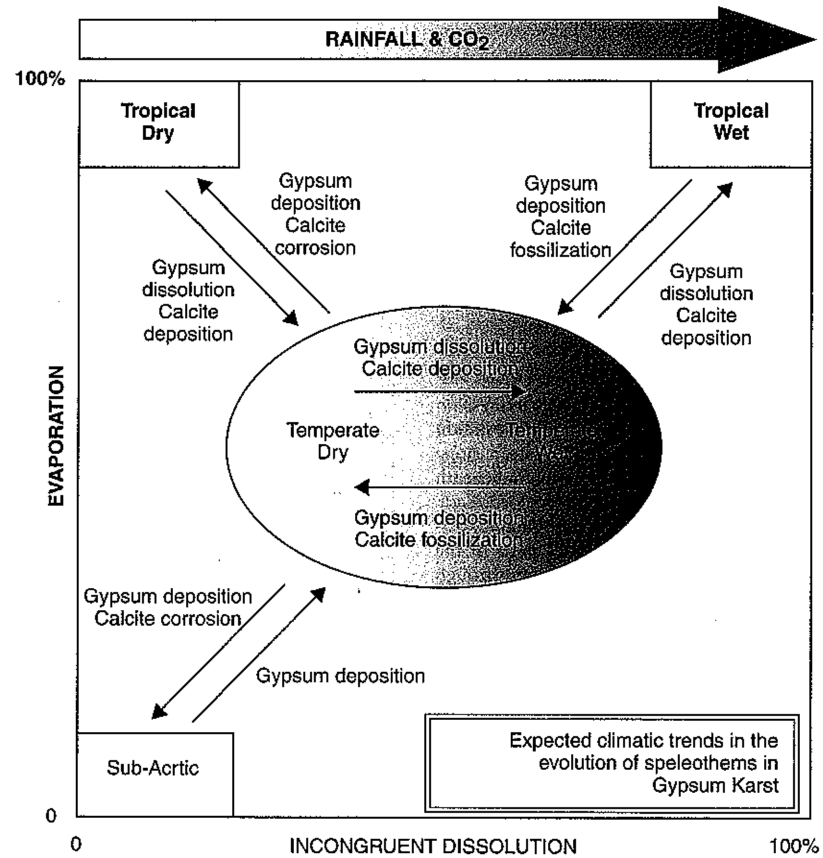


Figure 79 Expected climatic trends in the evolution of speleothems in gypsum caves

Cucchi, F., Forti, P. and Finocchiaro, F. (1998) Gypsum degradation in Italy with respect to climatic, textural and erosional conditions, in J. James. and P. Forti (eds) *Karst Geomorphology*, 41–49, Geografia Fisica e Dinamica Quaternaria supplement III, v.4.

Forti, P. (1996) Erosion rate, crystal size and exokarst microforms, in J.J. Fornos. and A. Gines (eds) *Karren Landforms*, 261–276, Universitat de Illes Balears, Mallorca.

Hill, C. and Forti, P. (1997) *Cave Minerals of the World*, Huntsville, AL: National Speleological Society.

Further reading

Klimchouk, A., Lowe, D., Cooper, A. and Sauro, U. (eds) (1996) Gypsum karst of the world, *International Journal of Speleology* 23(3–4).

PAOLO FORTI

H

HALDENHANG

Haldenhang is a German geomorphic expression introduced by W. Penck (1924, see translation 1953) for a 'basal slope – the less steep slope found at the foot of a rock wall, usually beneath an accumulation of talus'.

Figure 80 illustrates the formation of a haldenhang at the foot of a rock wall: all parts of the rock face but one are subject to erosion through rockfall, namely its base. The material there cannot fail because there is no gradient beneath it. This results in a parallel retreat of the rock face, during which its foot moves gradually upwards. Thus a rock slope of lower inclination appears, the haldenhang. Underneath a rapidly weathering rock face, the haldenhang may be covered by rockfall debris, forming a SCREE or TALUS. Over time the rock face will suffer incremental reduction in its height finally to be replaced by the haldenhang.

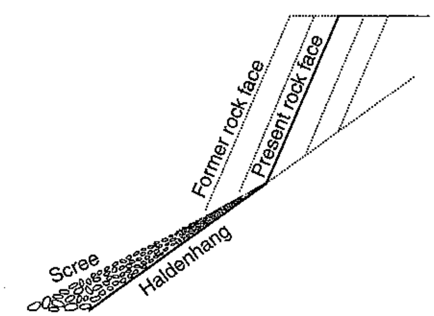


Figure 80 Formation of a haldenhang at the foot of a rock wall

W. Penck built his observations on haldenhang formation into his classic model of landform evolution and used them for explaining the erosional processes responsible for the downwearing of a relief (see SLOPE, EVOLUTION). According to Penck (1953), waning slope development starts with steep valley slopes being replaced by haldenhang of less inclination. Weathering of the surface material of the haldenhang will eventually produce finer material susceptible to creep and rainwash. Thus erosion of the haldenhang starts, resulting in a parallel retreat of the haldenhang and the development of a lower slope segment at its base. The whole process of slope retreat by mutual consumption produces concave and progressively lower slopes.

Reference

Penck, W. (1953) *Morphological Analysis of Land Forms*, trans. H. Czech and K.C. Boswell, London: Macmillian.

CHRISTINE EMBLETON-HAMANN

HAMADA

Large rocky, unvegetated plateaux spread over dozens of kilometres in the Sahara, the Australian deserts and Libya, e.g. Hamadas of Dra, and in the Guir, north-western Sahara (Mabbutt 1977).

The surface of hamadas shows STONE PAVEMENTs which could either be residual and result from the disintegration of rock formations below or consist of boulders transported only short distances, forming a reg. The reg acts as a protective layer to the underlying formations of the hamada which represent forms of stabilized relief. The hamadas of the Sahara rest on erosion surfaces of varied age: Cretaceous, Oligocene, Miocene, Pliocene or Quaternary (Conrad 1969). The hamadas of the northern Sahara are characterized by the extension onto the southern Atlas piedmont of the detrital and lacustrine facies of the *torba*, which is a non-stratified sediment, interrupted by one or several levels of silicified dolomitic limestones, called the *hamadienne carapace*. The geochemical interpretation of the torba and the carapace is that it formed by continental sedimentation in a lacustrine environment as indicated by the abundance of neoformed attapulgite, dolomite and calcite.

References

Conrad, G. (1969) *L'évolution continentale post-hercynienne du Sahara algérien (Saoura, Erg Chech, Tanezrouft, Ahnet-Mouydir)*, série Géologie No. 10, Paris: Centre National de la Recherche Scientifique, CNRS.

Mabbutt, J.A. (1977) *Desert Landforms, An Introduction to Systematic Geomorphology*, Cambridge, MA: MIT Press.

MOHAMED TAHAR BENAZZOUZ

HANGING VALLEY

A tributary valley in which the floor at the lower end is notably higher than the floor of the main valley in the area of junction. Hanging valleys are a hallmark of GLACIAL EROSION in mountains, because the greater bulk of the trunk glacier was able to cut a larger valley cross section than those of smaller tributary glaciers, in consequence of which the floor of the main valley was eroded to a lower level. The relationship between the size of valley glacier troughs and ice discharge was first noticed by A. Penck (1905) who termed it as the 'law of adjusted cross-sections'. Indeed, geomorphometric assessments undertaken in more recent years strongly support his idea that discharge and trough size are mutually adjusted (Benn and Evans 1998: 365).

Hanging valleys have a variety of forms. In high mountain areas the cross profile will exhibit the typical U-shape of glacial erosion. If the tributary valley was not glaciated or only occupied by thin, cold-based ice the preglacial V-shape might prevail. In some cases a waterfall cascades over the lip into the main valley, but in high mountain areas headward erosion of the tributary stream has usually cut a narrow gorge into the lower reaches of the hanging valley floor.

Outside previously glaciated areas hanging valleys sometimes occur along youthful fault scarps or along coasts where the rate of cliff retreat is higher than the adjustment potential of the smaller streams, e.g. in the chalk cliffs in the south of England. Hanging valleys can also develop in karst areas where surface streams flow directly on the groundwater surface. If the main river is downcutting rapidly, the water level will be progressively lowered and the smaller tributaries will eventually turn into DRY VALLEYs hanging above the main valley.

References

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.

Penck, A. (1905) Glacial features in the surface of the Alps, *Journal of Geology* 13, 1–17.

CHRISTINE EMBLETON-HAMANN

HEADWARD EROSION

Headward erosion is the process by which a stream extends upstream towards the catchment divide. Headward erosion occurs at a range of scales, from rills to large rivers, and in all environments. Drylands have been central to research into headward erosion because conditions that favour GULLY and ARROYO development are frequently found in these regions. Headward erosion may also be caused by RIVER CAPTURE.

In any channel network, approximately half of the total length of channels is in un-branched (first-order) fingertip tributaries. Environmental changes that promote channel extension therefore have a large potential impact on the landscape. During discharge events channel heads may advance great distances upslope, or retreat downslope if the hollow refills. In extreme cases, gullies can grow in length by tens of metres per year, and may also incise their channels creating steep ravine banks (Bull and Kirkby 2002). One possible end result of these processes is the creation of BADLANDs, where there is little or no remaining land suitable for agriculture.

Headward erosion occurs at the channel head (see CONTRIBUTING AREA). In terms of landscape dynamics, the channel head is one of the most important elements of the coupled hillslope-channel system. The location of the channel head controls the distance to the catchment divide and therefore influences the drainage density and average hillslope length of a catchment (although bifurcation frequency, confluence angles and tributary spacing are also important) (Bull and Kirkby 2002). The position of the channel head is controlled by the balance of sediment supply and sediment removal (Kirkby 1980; Dietrich and Dunne 1993). A change in any factor that influences this balance, such as fluctuations in climate or land use, alter the surface erodibility, sediment supply and runoff rates and may therefore result in headward erosion.

Channel extension results from a complex array of processes that reflect variations in slope, soil type, soil thickness, vegetation type and vegetation density. These processes include overland flow, pipe initiation and collapse, mass failures and hillslope processes (which have the reverse effect to headward erosion by infilling the channels).

Overland flow occurs in small rills or as sheets of moderate depth over large surfaces. For erosion to occur the rate of rainfall must be sufficient to produce runoff, and the shear stress produced by the moving water must exceed the resistance of the soil surface. Erodibility is a function of the permeability of the surface, the physical and chemical properties that determine the cohesiveness of the soil, and the vegetation.

In some areas there is a close association between piping (see PIPE AND PIPING and headward erosion. The erosive effects of flow through subsurface channels may result in TUNNEL EROSION and subsequent collapse to cause headward erosion. Piping intensity reflects a critical interaction between climate conditions, soil/regolith characteristics and local hydraulic gradients.

Mass failures also occur at channel heads to cause headward erosion. Failure of steep channel heads occurs when the driving forces exceed resisting forces. Channel heads are loaded by three different forces: (1) the weight of the soil, (2) the weight of water added by infiltration or a rise in the water table, and (3) seepage forces of percolating water (Bradford and Piest 1977). The change in water content is important because it has a strong influence on the shearing resistance of the soil. The shear strength is also influenced by freeze–thaw cycles and wetting–drying cycles.

Vertical tension cracks tend to decrease overall stability by reducing cohesion, and when these are filled with water the pore-water pressure increases dramatically, often resulting in failure.

Hillslope processes such as rainsplash, wetting and drying cycles and frost action operate to infill channels, and hence reverse headward erosion. For incisions to grow the rate of sediment transport out of an incision must also exceed the rate of sediment input at the same point, otherwise filling will occur. The inter-rill hillslope processes involved are rainsplash and rainflow. Both processes depend on raindrop impact to detach soil material. Mass failures may also act to fill channels if failed material is not removed, but builds up at the base of headcuts.

Traditionally there are two conceptual approaches to understanding processes operating at the channel head, the stability approach (Smith and Bretherton 1972) and the threshold approach (Horton 1945). The stability approach emphasizes that the channel head represents the point where sediment transport increases faster than linearly downslope. This usually requires wash processes to dominate. The threshold approach takes the view that the channel head represents a point at which processes not acting upslope become important. The balance of sediment still determines whether the channel head becomes stable or migrates, but changing process domains drive incision. However, it is not clear whether there is always a change in process at the headcut, or whether a change in the intensity of the process operating, or a variation in the spatial distribution causes incision. The different approaches tend to be better suited to different environments and determine the two extremes of a range of factors that combine to produce channel heads. These models assist our understanding and prediction of headward erosion.

References

Bradford, J.M. and Piest, R.F. (1977) Gully wall stability in Loess derived alluvium, *Journal of the American Soil Science Society* 41, 115–122.

Bull, L.J. and Kirkby, M.J. (eds) (2002) *Dryland Rivers: Hydrology and Geomorphology of Semi-Arid Channels*, Chichester: Wiley.

Dietrich, W.E. and Dunne, T. (1993) The channel head, in K. Beven and M.J. Kirkby (eds) *Channel Network Hydrology*, 175–219, London: Wiley.

Horton, R.E. (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, *American Geological Society Bulletin* 56, 275–370.

Kirkby, M.J. (1980) The stream head as a significant geomorphic threshold, in D.R. Coates and A.D. Vitek (eds) *Thresholds in Geomorphology*, 53–73, London: Allen and Unwin.

Smith, T.R. and Bretherton, F.P. (1972) Stability and the conservation of mass in drainage basin evolution, *Water Resources Research* 8, 1,506–1,529.

SEE ALSO: arroyo; badland; donga; gully; pipe and piping; tunnel erosion

LOUISE BRACKEN (NÉE BULL)

HIGH-ENERGY WINDOW

Neumann (1972) suggested that in the mid-Holocene on tropical coasts there was a period when wave energy was greater than now. This occurred during the phase when the present sea level was being first approached by the Flandrian (Holocene) transgression and prior to the protective development of coral reefs. The 'window' may have operated on a more local scale on individual reefs with waves breaking not on margins of an extensive reef flat as now, but more extensively over a shallowly submerged reef top prior to the development of the reef flat (Hopley 1984).

References

Hopley, D. (1984) The Holocene 'high energy window' on the central Great Barrier Reef, in B.G. Thom (ed.) *Coastal Geomorphology in Australia*, 135–150, Sydney: Academic Press.

Neumann, A.C. (1972) Quaternary sea level history of Bermuda and the Bahamas, *American Quaternary Association Second National Conference Abstracts*, 41–44.

A.S. GOUDIE

HILLSLOPE-CHANNEL COUPLING

Fluxes from hillslopes to the channel system are controlled by the connectivity of process domains between different elements of the catchment system. Brunsden (1993) defines coupled systems as being ones where there is a free transmission of energy between elements, for example where a river channel directly undercuts a hillslope, whereas decoupled systems are ones where a barrier is present, for example in the case of a FLOODPLAIN buffering the input of the hillslope to the channel. The extent to which a hillslope is coupled to the channel is thus a function of any factor that affects its connectivity, and may relate to spatial variability of properties such as soil texture or vegetation cover (see OVERLAND FLOW). A floodplain may cause a hillslope to be strongly coupled to the channel if it has a low enough infiltration rate, or at times when it is already saturated. The main channel may be decoupled from the hillslope by the presence of minor channels running along the edge of floodplains (YAZOO channels) or human-made drainage channels.

The strength of coupling may affect the type of process that occurs on either side of the boundary. A channel directly undercutting a hillslope may cause the local gradient to be steep enough to initiate RILLS or gullies (see GULLY) on the hillslope, or may lead to failure of the base of the slope (e.g. Harvey 1994). In all cases, the amount of sediment fed into the channel will increase, and may cause it to avulse (see AVULSION) or change its planform (see BRAIDED RIVERS). The rate of removal of sediment from the base of a hillslope relative to its supply by processes on the slope will also affect the form of slope evolution (see SLOPE, EVOLUTION) in the longer term. Coupled slopes will tend to have more convex lower profiles whereas decoupled slopes will encourage deposition at the slope base leading to concave lower profiles. Strongly coupled slopes will also be more sensitive to changes elsewhere in the catchment system.

Consideration of the strength of coupling may also be important in an APPLIED GEOMORPHOLOGY context. SLOPE STABILITY from undercutting is again an important process here, while Burt and Haycock (1993) discuss the impact of floodplain buffers on water quality, for example due to pollutants carried by runoff (see RUNOFF GENERATION) from hillslopes.

References

Brunsden, D. (1993) The persistence of landforms, *Zeitschrift für Geomorphologie, Supplementband* 93, 13–28.

Burt, T.P. and Haycock, N.E. (1993) The sensitivity of rivers to nitrate leaching: the effectiveness of near-stream land as a nutrient retention zone, in D.S.G. Thomas and R.J. Allison (eds) *Landscape Sensitivity*, 261–272, Chichester: Wiley.

Harvey, A.M. (1994) Influence of slope/stream coupling on process interactions on eroding gully slopes, in M.J. Kirkby (ed.) *Process Models and Theoretical Geomorphology*, 247–270, Chichester: Wiley.

Further reading

Harvey, A.M. (2002) Effective timescales of coupling within fluvial systems, *Geomorphology* 44, 175–201.

Michaelides, K. and Wainwright, J. (2002) Modelling the effects of hillslope-channel coupling on catchment hydrological response, *Earth Surface Processes and Landforms* 27, 1,441–1,457.

JOHN WAINWRIGHT AND KATERINA MICHAELIDES

HILLSLOPE, FORM

What are hillslopes?

Most of the Earth's surface is occupied by hillslopes. Hillslopes therefore constitute a basic element of all landscapes (Finlayson and Statham 1980) and a fundamental component of geomorphologic systems (see SYSTEMS IN GEOMORPHOLOGY). However, there is an 'amazing absence of any precise definition' of hillslopes (Schumm and Mosley 1973; Dehn *et al.* 2001). Hillslopes have a very large variety of sizes and forms; and several more or less synonymous terms are used to describe the phenomenon hillslope, e.g. valley slope, hillside slope, mountain flank. The description of hillslope form is a fundamental problem in geomorphology (see GEOMORPHOMETRY).

Generally, a hillslope is a landform unit, that is, a part of the Earth's surface, with specific characteristics (see LAND SYSTEM). As a basic characterization, a hillslope can be defined as an *inclined landform unit* with a slope angle larger than a lower threshold β_{min} (delimiting hillslopes from plains) and smaller than a higher threshold β_{max} (delimiting hillslopes from vertical walls like cliffs or overhangs), which is *limited by an upper and a lower landform unit* (Dehn *et al.* 2001). A definition of hillslopes additionally has to include position within the landscape as an *external context.* A valley, for example, can only exist with its accompanying hillslopes. Moreover, *size and scale* context are important properties for definitions of hillslopes: a hiker in Grand Canyon might identify the components of the valley side as an individual hillslope itself, whereas a pilot flying over the scene defines the whole canyon side as a hillslope. Hillslopes are formed as the result of hillslope processes (see HILLSLOPE PROCESS) acting over different timescales. Therefore, hillslopes are units, where downslope component of gravitational stress (g sin β) plays a dominant role for acting hydrologic and geomorphologic processes. However, a hillslope is usually the product of a variety of processes interacting in space and time; therefore, hillslopes form sequences of hillslope units with different characteristics (compare Figure 81; see SLOPE, EVOLUTION).

Therefore, fundamental properties for the definition of a hillslope are: (1) local geometry, (2) external landform relationships, (3) scale, and (4) related processes. The utilization of these fundamental properties into a definition of hillslope depends on the perception or the specific application, that is, a specific semantic model for hillslopes (Dehn *et al.* 2001). In geomorphometry, hillslope forms are usually described as arrangements of individual hillslope units. This concept facilitates description and classification of hillslopes, and enables modelling of interaction of hillslope form with forming geomorphologic processes. Hillslope analysis therefore incorporates two major connected aspects: the decomposition of a hillslope profile into units, and the aggregation of a hillslope by arrangements of form units. These analysis steps are carried out in three dimensions for a hillslope or, in a simplified way, two dimensionally for a hillslope profile. The related terminology used here is listed in Table 23.

Hillslope units

Hillslope analysis is carried out by subdivision of a hillslope into different hillslope units. There have been several approaches to standardize hillslope units using qualitative terms (e.g. Speight 1990). Most commonly, a hillslope is described by a series of basic units describing changes in slope, curvature and processes along the hillslope profile.

- Ridge/crest/interfluve: convex/rectilinear unit; most stable unit in landscape, if of considerable width; mainly vertical water transport; more poorly drained soils.
- Shoulder/upper midslope: convex element; unstable unit due to erosion processes; minimum soil thickness.
- Backslope/midslope: usually rectilinear segment; unstable unit; intensive lateral drainage; sediment transport; soils of varying depth.
- Footslope/lower midslope: concave element; sediment deposition; unstable unit; soil thickness tends to increase.
- Toeslope/floodplain: concave/rectilinear unit; sediment input from upstream and hillslope; unstable unit; thicker soils.

Table 23 Basic components and terminology for hillslope analysis

Hillslope component	Definition
Hillslope profile	Flowline connecting drainage divide with thalweg
Hillslope toposequence	Arrangement of hillslope units within the hillslope
(Hill)slope unit	Part of hillslope with specific characteristics: segment or element
Segment	Unit of homogeneous slope angle
Element	Unit of homogeneous curvature
Convex element	Element with a downslope increase in angle
Concave element	Element with a downslope decrease in angle
Maximum segment	Segment, steeper than units above and below
Minimum segment	Segment, gentler than units above and below
Crest segment	Segment bounded by downward slopes in opposite directions
Basal segment	Segment bounded by upward slopes in opposite directions
Irregular unit	Slope unit with frequent changes of both angle and curvature

Source: Young (1972, modified and extended)

In the direction of contours, hillslopes are usually stratified into the elements HILLSLOPE HOLLOWS, spurs (or noses), and rectilinear valley sideslopes using plan curvature. Another method for hillslope unit classification is based on the position within the DRAINAGE BASIN: Young (1972: 4) distinguishes the 'component slopes': valley-head slopes, spur-end slopes and valley side slopes. Speight (1990) provided an exhaustive list of nomenclatures of different land elements, including many hillslope units. Hillslope units therefore generally incorporate different aspects of land surface form: (1) slope angle, (2) curvature, (3) position within drainage basin and (4) position within hillslope. These properties are used to derive quantitative models of hillslope units (see below). Moreover, hillslope units are related to different geomorphic processes and REGOLITH properties (see above, compare Speight 1990). This leads to the utilization of hillslope units for soil-landscape modelling (see SOIL GEOMORPHOLOGY), formalized for example in the concept of the CATENA.

One of the early quantitative approaches in hillslope analysis, based entirely on geomorphometric properties, was established by Savigear (1952) using the components profile intercept (constant slope gradient), slope segment (constant slope gradient consisting of several profile intercepts), and slope element (constant convex or concave curvature). These units are delimited by breaks of slope, which are characterized by a distinct change in slope gradient. This approach has been extended and quantified by Young (1972), who subdivided the slope into convex, concave and rectilinear units (Table 23).

Those early approaches mostly concentrated on quantitative description of hillslope profiles; however, a hillslope is not simply a linear feature, but a two-dimensional landform unit within the three-dimensional space, acting as a boundary layer of a three-dimensional lithological body. Characterization of local hillslope form is therefore based on the land surface derivatives: gradient, which has two components, slope angle and aspect angle; and curvature, which is usually described by two components in profile and contour or tangential directions (see MORPHOMETRIC PROPERTIES). Curvature can be classified into convex, concave and rectilinear surfaces (Young 1972). Hence, the combination of three slope profile curvature characteristics and three plan curvature characteristics leads to nine possible hillslope units, which are defined by Dikau (1989) as basic form elements of the landscape (Figure 81). They deliver a disjunctive description of the hillslope surface into units of homogeneous curvature characteristics.

Slope angle has been used to describe hillslopes by slope segments (Table 23). Young (1972: 173) compared several classifications of slope angle and proposed a system of seven classes: 0°–2° level to very gentle; 2°–5° gentle; 5°–10° moderate; 10°–18° moderately steep; 18°–30° steep; 30°–45° very steep; >45° precipitous to vertical/overhanging. Limiting angles describe the range of slope

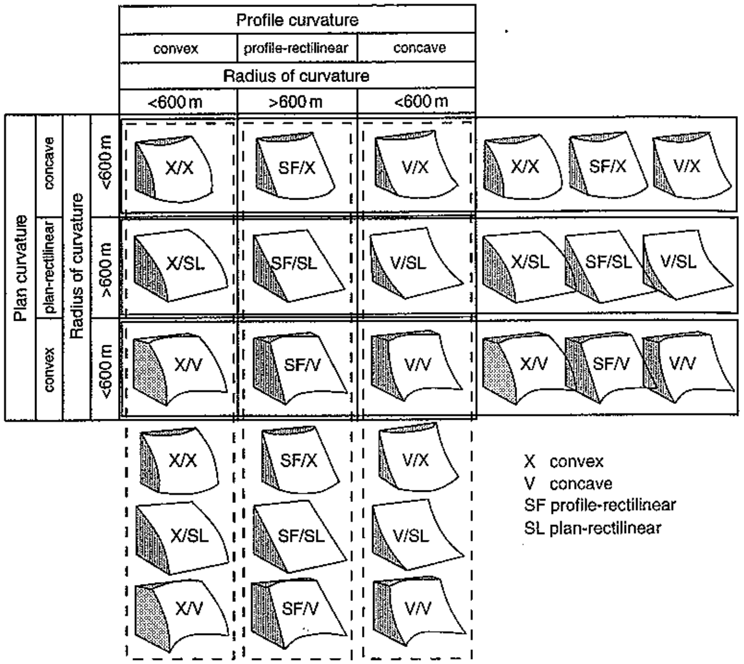


Figure 81 Fundamental hillslope form elements classified by plan and profile curvature (Dikau 1989)

angles within which specific slope forms occur. They include maximum and minimum limiting angles, which are related to the environmental conditions and the corresponding geomorphic processes. The angle of repose (see RESPOSE, ANGLE OF) defines the maximum angle for a given granular material type. Young (1972: 165) gives some figures for limiting angles of hillslope units under different environmental and lithological conditions.

An additional parameter for description of hillslope units is the position in the toposequence (Table 23) of the hillslope. Usually qualitative terms as upper slope, mid-slope, and lower slope are applied. Young (1972) used neighbourhood relationships to upper/lower units to describe hillslope units (see e.g. maximum and minimum segment in Table 23). However, for more complex profiles, these measures fail to describe absolute hillslope position, and quantitative rules, derived from total hillslope length or height, need to be introduced.

Investigations on hillslope forms are centred on process-form relationships, i.e. the explanation of a specific slope form (slope unit) by hillslope evolution, and contemporary processes. Young (1972: 92) gives a series of classical explanations of convex, concave and rectilinear slopes. Convex slope elements generally indicate erosion processes, which increase with slope length (surface wash), additionally soil creep and weathering has been identified as the dominating process regime for convex slopes. Rectilinear slopes generally indicate spatially homogeneous erosion conditions, i.e. the slope retreats parallel, or static transport units. Concave parts of hillslopes are explained by sediment accumulation due to constant base levels and/or by surface wash as an analogue to graded river profiles. However, as hillslopes are complex phenomena with an evolutionary history over long timescales, many interactions of processes and components can occur. Therefore, such simple assumptions generally do not match with the specific hillslope case and can only be used as guidelines.

Hillslope profiles and toposequences

Complete hillslopes are often represented by *hillslope profiles*. According to Young (1972) and Parsons (1988) a hillslope profile can be defined as a line on a land surface, connecting a starting point at the drainage divide with an end point at the thalweg, following the direction of the steepest slope. Hillslope profiles have been used to characterize various types of terrain using typical distribution of slope angles. Differences in the frequency distributions of slope angle are related to lithology (material resistance), climate (stress through rainfall, temperature), and evolutionary state of the slope (see limiting angles above). Hillslope profiles usually cover several process domains. Often the upper section of a slope is characterized by erosion, the middle section by transport and the basal section by deposition. Therefore, *toposequences* are used to describe the arrangement of different units within the hillslope profile. Characterization of these sequences delivers information about the slope system. It can be used to classify hillslopes. A toposequence may include one simple slope (single-sequence, e.g. a rectilinear element connecting ridge and valley), or two or more units (multi-sequence, e.g. convexo-concave slopes) (Speight 1990: 14). For multi-sequences, the order of the slope units (e.g. XMV for convex–rectilinear–concave slope) and the proportional length of the same units can be used to characterize the whole profile (Young 1972: 189).

The description of slope units within the context of the entire two-dimensional slope in three-dimensional space is also part of a toposequential analysis. Dikau (1989) and Schmidt and Dikau (1999) used parameters such as the neighbourhood relationship, distance to drainage divide, or height difference to the drainage channel to classify and aggregate complex hillslope systems.

Different models of hillslope profile form have been developed, which relate a specific hillslope toposequence to evolutionary history and contemporary geomorphic processes. Wood (1942) introduced the term 'waxing slopes' for convex hillslope units on crests developed by weathering processes acting on a cliff top. Likewise, Wood defines 'waning slopes' as depositional concave hillslope units developing at the base of a SCREE by sediment sorting due to aquatic hillslope processes.

King (see Young 1972: 37) developed a classical four-unit toposequential model based on work of Wood (1942). The crest (waxing slope) is a convex element of little erosion by weathering and creep processes. The evolution of the whole hillslope is driven by an active scarp segment of steep slope angle (rill erosion, mass movements). The downslope debris slope segment is formed by sediment provided by the scarp and determined by the angle of repose of the coarser material. The PEDIMENT (waning slope) is a rectilinear-concave, upward erosional element, produced by surface wash that connects to alluvial plains. Dalrymple *et al.* (1969) developed King's toposequence into a nine-unit slope model (Figure 82). The sequence consists of three low erosional upslope units, an intensively erosional unit (4), a transformational midslope (5), the depositional colluvial footslope (6) and three low-angled units associated with fluvial work. Conceptual hillslope models like these contribute to an understanding of the function of hillslope units and hillslope sequences, and can be utilized to classify hillslopes according to dominant process regimes. Numerical hillslope models (see MODELS) are used today to simulate process behaviour and evolution predicted by conceptual hillslope models, and thereby contribute to understanding of hillslope form and evolution based on current knowledge of process physics.

Measurement and analysis of hillslopes

There exist numerous techniques to measure hillslope froms. A selection of direct manual methods based on field observations and indirect measurements from maps and aerial photographs are described in Goudie (1990). Advances in computer technologies and the availability of DIGITAL ELEVATION MODELS (DEMs) have significantly revolutionized hillslope form analysis in the last decades. GIS technologies (see GIS) with algorithms to calculate morphometric properties, including slope gradient, slope curvature and flow paths, are now common tools for geomorphometric hillslope analysis (Schmidt and Dikau 1999). Raster-based DEMs are available at increasingly higher resolutions and, through the development of satellite data, on a global extent. As a result many numerical geomorphometric applications are raster-based. A typical GIS for hillslope analysis includes the following components (Table 24). Local morphometric properties of hillslopes

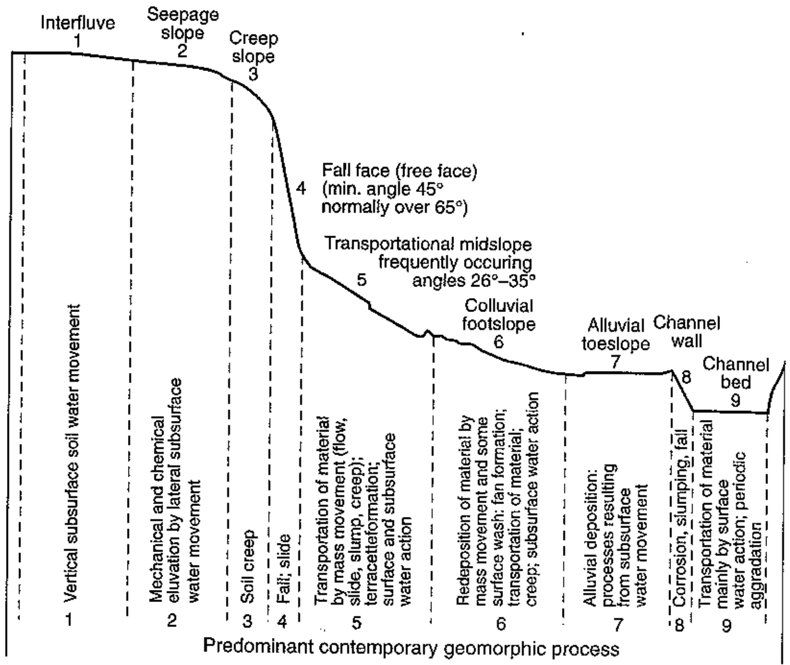


Figure 82 The nine-unit slope model of Dalrymple *et al.* (1969) (modified)

(height, curvature, gradient; see Schmidt and Dikau 1999) are generated from gridded DEMs through local interpolation, whereas complex parameters are based on flow routing algorithms. As slope and curvature are strongly dependent on scale, effects of DEM resolution on these parameters have to be considered, and preferably a specific scale for calculation of derivatives should be chosen. Hillslope units (areal geomorphometric objects after Schmidt and Dikau 1999) can be derived by GIS-based classification of slope, curvature and hillslope position (Dikau 1989). Linear hillslope profiles can be derived directly from a DEM by flow routing algorithms (Rasemann *et al.* 2003). Hillslopes as toposequences (geomorphometric objects of higher level after Schmidt and Dikau 1999) are derived by combining hillslope units according to their properties in gradient, curvature, position, and neighbourhood relationships. Hillslopes are characterized using representative geomorphometric parameters (Schmidt and Dikau 1999): frequency distributions and statistical moments of slope angle have been used to describe different types of slope profiles (Young 1972; Schumm and Mosley 1973). However, modern GIS technologies allow the calculation of many more hillslope parameters, including toposequential characteristics (Schmidt and Dikau 1999).

References

Dalrymple, J.B., Blong, R.J. and Conacher, A.J. (1969) A hypothetical nine-unit landsurface model, *Zeitschrift für Geomorphologie* 12, 60–76.
Dehn, M., Gärtner, H. and Dikau, R. (2001) Principles of semantic modeling of landform structures, *Computers and Geosciences* 27, 1,005–1,010.
Dikau, R. (1989) The application of a digital relief model to landform analysis in geomorphology, in J. Raper (ed.) *Three-dimensional Applications in Geographical Information Systems*, 51–77, London: Taylor and Francis.
Finlayson, B. and Statham, I. (1980) *Hillslope Analysis*, London: Butterworth.

Table 24 Hillslope analysis in a GIS

Hillslope parameters and objects	Algorithm	Input
Local geomorphometric parameters – local geometry		
Slope angle and aspect, profile and plan curvature	Local interpolator	DEM
Complex primary geomorphometric parameters – hillslope position		
Flowdirection	Local classifier	DEM
Upslope contributing area	Flow routing	Flowdirection
Downslope/upslope flowlength	Flow routing	Flowdirection
Hillslope position	Algebraic operation	Flowlength
Geomorphometric objects – hillslope units		
Form elements	Classification	Curvature
Slope segments	Classification	Slope angle
Hillslope units	Classification	Hillslope position
Geomorphometric objects – hillslopes		
Hillslope profile/flowpath	Flow routing	Flowdirection
Hillslope toposequence	Neighbourhood analysis	Form elements, slope segments
Representative parameters – hillslope characteristics		
Frequency distribution	Grid overlay	Hillslope profiles, Hillslope toposequences
Statistical moments of parameters	Grid overlay	Hillslope profiles, Hillslope toposequences

Notes: Local geomorphometric parameters are derived through a local interpolator. Complex geomorphometric parameters, which are related to landscape position, are derived through flow routing. Hillslope units and hillslope profiles are derived as geomorphometric objects. Hillslopes as a toposequential arrangement of units can be analysed by topological relationships of hillslope units

Goudie, A. (ed.) (1990) *Geomorphological Techniques*, London: Unwin Hyman.
Parsons, A.J. (1988) *Hillslope Form*, London: Routledge.
Rasemann, S., Schmidt, J., Schrott, L. and Dikau, R. (2003) Geomorphometry in mountain terrain, in M.P. Bishop and J.F. Schroder (eds) *Geographic Information Science (GIScience) and Mountain Geomorphology*, Berlin: Praxis Scientific Publishing.
Savigear, R.A.G. (1952) Some observations on slope development in South Wales, *Transactions Institute of British Geographers* 18, 31–51.
Schmidt, J. and Dikau, R. (1999) Extracting geomorphometric attributes and objects from digital elevation models – semantics, methods and future needs, in R. Dikau and H. Saurer (eds) *GIS for Earth Surface Systems*, 153–173, Stuttgart: Schweizerbarth.
Schumm, S.A. and Mosley, M.P. (eds) (1973) *Slope Morphology. Benchmark Papers in Geology*, Stroudsburg, PA: Dowden, Hutchinson and Ross.
Speight, J.G. (1990) Landform, in R.C. McDonald, R.F. Isbell, J.G. Speight and J. Walker (eds) *Australian Soil and Land Survey: Field Handbook*, 9–57, Melbourne: Inkata Press.
Young, A. (1972) *Slopes*, London: Longman.
Wood, E.B. (1942) The development of hillside slopes, *Proceedings of the Geological Association* 53, 128–140.

Further reading

Ahnert, F. (1970) An approach towards a descriptive classification of slopes, *Zeitschrift für Geomorphologie, N.F. Supplementband* 9, 71–84.
Carson, M.A. and Kirkby, M.J. (eds) (1972) *Hillslope – Form and Process*, Cambridge: Cambridge University Press.
Pike, R. and Dikau, R. (eds) (1995) Advances in geomorphometry, *Zeitschrift für Geomorphologie N.F. Supplementband* 101, Stuttgart: Schweizerbarth.

SEE ALSO: catena; cliff, coastal; digital elevation model; drainage basin; geomorphometry; GIS; hillslope hollow; hillslope, process; morphometric properties; pediment; repose, angle of; slope, evolution; valley.

RICHARD DIKAU, STEFAN RASEMANN AND JOCHEN SCHMIDT

HILLSLOPE HOLLOW

Hillslope hollows are elongate depressions within the bedrock of regolith mantled hillslopes. They have no obvious stream channel but serve as drainage lines that are integrated with the drainage network by either subsurface or surface topography. They belong to a morphological spectrum ranging from small low-relief upland depressions, a few tens of metres in length, through dells which may extend 200–300 m in length (Ahnert 1998), to valley head depressions, and DRY VALLEYs which have no active channels but are clearly of fluvial origin. Such features are found in many different countries, in a wide range of morphoclimatic and lithological conditions, and have many different modes of origin; thus giving rise to a varied and at times contradictory terminology. For example, United Kingdom usage represents a 'dell' as 'a small well-wooded stream or river valley' while in many other countries a dell is defined as 'a small dry valley with no trace of linear, fluvial erosion' and of periglacial origin (Fairbridge 1968: 250). Other forms of hollow occurring on hillslopes that are not integrated with the drainage network (e.g. individual landslide scars or fault sags) are excluded from this discussion. Reneau and Dietrich (1987) provide a literature review on hillslope hollows.

Hillslope hollows within the substrate are often filled with colluvium and other regolith material (Plate 60) and may or may not exhibit a depression in the landsurface. Because they can lack surface expression, the term 'colluvium-filled bedrock depression' has been suggested as a more accurate descriptive term for these features. In one study of 80 hillslope hollows exposed in road cuts, 37 were found to be associated with concave depressions in the slope surface, 35 occurred beneath planar slopes with no surface depression and 8 occurred on spurs. The cross-sectional form of the depression within the bedrock can be either V-shaped or broad-based (Crozier *et al.* 1990). Hillslope hollows are located headward of first-order channels or join higher order channels in positions similar to that occupied by first-order channels. Tsukamoto (1973) has used the term 'zero-order basin' to describe landsurface hollows, emphasizing their hydrological integration with the drainage network.

Plate 60 V-shaped colluvium-filled bedrock depression, in Pliocene marine sediment, Taranaki, New Zealand

The configuration of hillslope hollows in valley head settings (Figure 83) can be related to contemporary processes (Ahnert 1998; Montgomery and Dietrich 1989). The main criteria used to differentiate the various valley head forms are gradient, number of convergent hollows, and shape. There are four basic types that can be further subdivided on the basis of number of contributing hollows. Shallow gentle hollows (Figure 83a) are wide with a low gradient head and commonly a downslope topographic threshold. During prolonged rainstorms, this form of hollow generally produces saturated overland flow. Steep narrow hollows (Figure 83b), often with a head cut in the colluvial fill, are dominated by seepage and regolith landsliding. Funnel-shape valley heads (Figure 83c) usually result from the convergence of multiple steep hollows. Spring-sapping valley heads (Figure 83d) tend to be circular in shape with a distinctive low angle floor separated from convergent hollows by a marked break in slope at the spring line.

Several processes have been suggested as capable of producing hillslope hollows in different geomorphic settings (see Crozier *et al.* 1990), including landsliding, subaerial fluvial dissection, subsoil percoline erosion, gelifluction and seasonal meltwater and a combination of periglacial and fluvial processes. Cotton and Te Punga, (1955) demonstrate that hillslope hollows are products of alternating morphoclimatic regimes. They conclude that former stream channels, initially incised during the last interglacial became modified by periglacial mass movement processes in the subsequent glacial episode and eventually infilled by periglacial deposits and loess. Under present-day conditions these colluvial infills are being removed by shallow landsliding. Another cyclic infilling and evacuation process, but in this case involving landsliding as the initial hollow generating process (Figure 84),

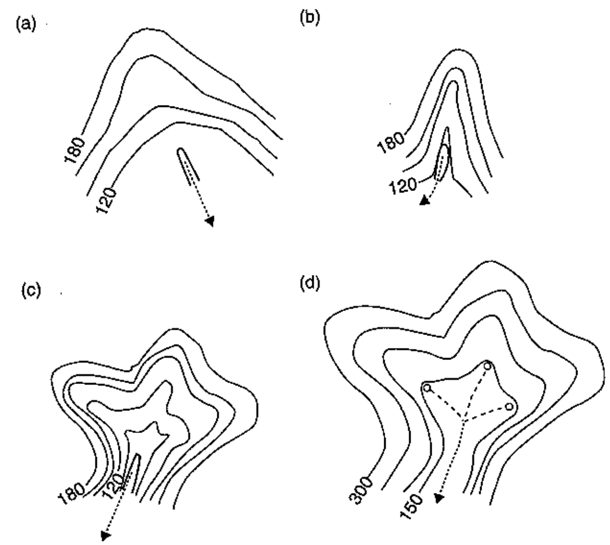


Figure 83 Contour pattern for types of valley head hollows: (a) shallow gentle hollow; (b) steep narrow hollow; (c) funnel-shaped valley head with three convergent hollows; (d) spring-sapping valley head (based on descriptions of Ahnert (1998) and Montgomery and Dietrich (1989))

has been described by Dietrich and Dunne (1978).

During rainstorms, the geometry of hillslope hollows directs both surface and subsurface runoff towards the centre line of the hollow. Accumulation of water within the hollow is a function of the contributing area and the ratio of side slope to thalweg gradients. During prolonged rainstorm events the hollows may become preferentially saturated acting as a source area for saturated overland flow. Regolith-filled hollows are also a preferential location for the generation of debris flows and debris slides. Compared to other hillslope locations, perched water tables are more readily established and accumulation of sediment surpasses the critical thickness required for failure. The scale of landsliding that occurs under these conditions is a function of the number of convergent hillslope hollows and their volume of stored sediment.

References

Ahnert, F. (1998) *Introduction to Geomorphology*, London: Arnold.

Cotton, C.A. and Te Punga, M.T. (1955) Solifluxion and periglacially modified landforms of Wellington, New Zealand, *Transactions Royal Society, New Zealand* 82(5), 1,001–1,031.

Crozier, M.J., Vaughan, E.E. and Tippett, J.M. (1990) Relative instability of colluvium-filled bedrock depressions, *Earth Surface Processes and Landforms* 15, 329–339.

Dietrich, W.E. and Dunne, T. (1978) Sediment budget for a small catchment in mountainous terrain, *Zeitschrift für Geomorphologie Supplementband* 29, 191–206.

Fairbridge, R.W. (1968) Dell, in R.W. Fairbridge (ed.) *The Encyclopedia of Geomorphology*, 250–252, New York: Reinhold.

Montgomery, D.R. and Dietrich, W.E. (1989) Source areas, drainage density and channel initiation, *Water Resources Research* 26, 1,907–1,918.

Reneau, S.L. and Dietrich, W.E. (1987) The importance of hollows in debris flow studies: examples from Marin County, California, in J.E. Costa and G.F. Wieczorek (eds) *Debris flows/Avalanches; Processes, Recognition, and Mitigation*, Geological Society of America, *Reviews in Engineering Geology* 7, 1–26.

Tsukamoto, Y. (1973) Study on the growth of stream channel, (1) Relation between stream channel growth and landslides occurring during heavy rainstorm, *Shin-Sabo* 25, 4–13.

MICHAEL J. CROZIER

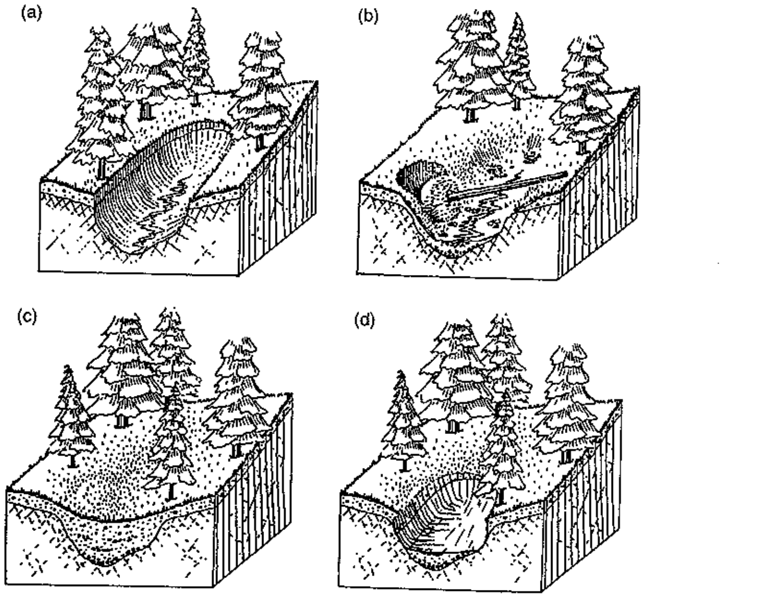


Figure 84 A model for the origin and evolution of hillslope hollows (based on Dietrich and Dunne (1978)): (a) bedrock landslide produces initial hollow; (b) peripheral debris fills hollow and is sorted by fluvial processes; (c) filled hollow becomes a site of concentrated subsurface flow and potential debris slide; (d) evacuation by debris slide

HILLSLOPE, PROCESS

The form that hillslopes take is a product of the materials of which they are made and the forces that act upon them. While the gross form of the landscape is determined by many factors, at the local scale a range of characteristic geomorphic processes shapes hillslopes. At the same time, hillslope morphology acts as a major influence on the occurrence, magnitude and nature of the processes themselves. Thus form and process tend toward a mutual adjustment. Hillslope processes are those geomorphic processes that involve the entrainment, transport and deposition of material from, over and on slopes. Their net effect is the transfer of material to lower parts of the landscape. This occurs either under the influence of gravity alone or, more commonly, with the additional incorporation of varying amounts of water. Flowing ice and wind also contribute to hillslope form, but operate at scales that are greater than the hillslope (see GLACIFLUVIAL; AEOLIAN PROCESSES; WIND EROSION OF SOIL).

A useful distinction can be drawn between two sets of processes that differ with respect to the role that water plays in the entrainment phase. First, MASS MOVEMENTs involve the entrainment of material due to the effect of gravity. The impetus for movement of material is derived from the potential energy inherent in the material by virtue of its elevation above BASE LEVEL, and the magnitude of the potential energy gradient induced by the inclination of both hillslope surface and strata. Water may play a crucial role as a preparatory or triggering factor, but it is the balance of geomechanical stresses within the mass of material that determines whether entrainment occurs (see FACTOR OF SAFETY). Once failure has occurred water is also significant in determining the nature of subsequent transport. In a second set of hillslope processes both entrainment and transport are effected directly by the kinetic energy of moving water.

The term mass movement can be applied to a broad spectrum of processes that are often classified in terms of the type of material involved (e.g. bedrock, debris or earth) and the type of movement (e.g. fall, topple, slide, spread and flow). This range of types of movement reflects an increasing significance of water in the transport phase; falls are almost exclusively gravitational, while, at the other extreme, flows almost always require the presence of water. A similar spectrum exists with respect to the importance of water for entrainment of material, i.e. the initiation of mass movement. In some instances failure may occur simply due to the effects of gravity, e.g. a rockfall. More commonly, water will have some influence on the balance of stresses within the slope material. Water is an important factor in preparing soil and regolith material for mass movement, and can often also be a triggering factor. The effects of water in various forms of physico-chemical WEATHERING produce a slow reduction in shear strength as rock is transformed into an engineering soil and rendered increasingly susceptible to gravitational stresses. More dynamically, the behaviour of water within hillslope materials frequently triggers failure. Variation in the height of groundwater tables can alter the balance of stresses by both increasing the mass of material (increased shear stress) and reducing its shear strength as a result of elevated PORE-WATER PRESSURES or reduced COHESION.

The second set of processes involves the entrainment of material through the energy imparted by flowing or impacting water. While not the direct cause of entrainment, gravity plays an important role in determining the velocity of flow, and hence its energy, and the direction of transport. The energy of a raindrop impact (see RAINDROP IMPACT, SPLASH AND WASH) is capable of detaching soil particles, which may be transported by the resulting splash. Although splash may be in all directions, the net effect is of downslope transport. Greater volumes of material may be entrained and transported by water flowing over the surface of unprotected soils (see SHEET EROSION, SHEET FLOW, SHEET WASH; OVERLAND FLOW). The mass of material that can be entrained will be controlled by the shear stress that is applied to the soil or regolith surface, which will be determined by the relationship between the HYDRAULIC GEOMETRY of the flow and the ROUGHNESS of the surface. The distance over which material of a given size can be transported is proportional to the kinetic energy of the flow. Where microtopographic convergences induce concentration into linear and turbulent flow, a threshold (see THRESHOLD, GEOMORPHIC) is reached and RILLs are formed. Once initiated, rills act to further concentrate flow. This concentrated flow can result in the formation of a GULLY, giving an example of positive feedback. Another important process involving flowing water is TUNNEL EROSION, which occurs when subsurface flow of water through the soil/regolith matrix is of sufficient velocity to initiate particle entrainment (see PIPE AND PIPING).

Although useful to illustrate the differing roles that water may play in hillslope processes, the distinction between mass movement and aquatic processes is often indistinct. The boundary between a highly fluid mass movement (e.g. an earthflow) and a sediment-laden stream may not always be easily identified. Strictly speaking, a rheologic distinction can be made between Newtonian and non-Newtonian flows; the former will involve both vertical and horizontal sorting of clasts during transport, while the latter implies matrix support and an absence of sorting. In reality the removal of material from slopes will occur as a result of a complex interaction between a range of different processes. A more important distinction can be made between those geomorphic processes that are diffuse and those that are not.

Diffusivity

Diffusivity is the property of being spread or dispersed, of not being specifically associated with any one place. Diffusive geomorphic processes, therefore, are those that are widely distributed in space. More importantly, diffusivity refers to the dissipation of energy. Thus, diffusive geomorphic processes can also be defined as those that occur where energy is dissipated over large areas (e.g. SOIL CREEP, sheet wash), while linear processes are characterized by the concentration of energy in a discrete unit of space (e.g. DEBRIS FLOW, RILLs). Consistent with the MAGNITUDE–FREQUENCY CONCEPT, low energy geomorphic processes occur frequently and with a wide spatial distribution. Conversely, high magnitude processes – those that concentrate large amounts of energy – occur more rarely and in a limited number of places. Hence, although diffuse geomorphic processes are often unspectacular, over longer periods they can accomplish large amounts of geomorphic work. Indeed, the predominance of either diffusive or non-diffusive processes has a large bearing on hillslope form (see HILLSLOPE, FORM).

Diffusivity tends to produce convex hillslope profiles. This is because of the adjustment of form to the energy available for entrainment and transport of material. On upper slopes close to drainage divides, with minimal catchment area, energy remains dissipated. Only small amounts of material can be transported. As catchment area increases with increasing distance from the divide, so too does available energy, and thus the amount of material that can be transported. The net effect of increasing entrainment and removal with greater distance from the drainage divide is the development of a convex long profile. At some distance from the drainage divide, available energy will be sufficient for the initiation of concentrated processes. This represents a threshold between diffusive and non-diffusive processes. It also represents a morphological threshold, and below this point long profiles are typically concave.

Characteristic form is therefore a reflection of process domain, and characteristic hillslope forms can be illustrated with reference to the processes that formed them. This may be on a local scale, with diffusivity implying that different processes are predominant on different parts of hillslopes (see, for example, the nine unit hillslope model developed by Dalrymple *et al.* (1969)). Available energy determines whether diffuse or concentrated processes can occur, and therefore how much geomorphic work can be done. There is thus a zone in which diffuse processes, accomplishing smaller amounts of work, predominate. These zones tend to remain constant; diffuse processes produce a characteristic low energy form, which further determines that only these low energy processes occur. Similarly, high energy processes tend to maintain the form that is necessary for their initiation until there is no longer sufficient potential energy available for their initiation (see PENEPLAIN).

However, there is a temporal element in this distinction. Given a sufficiently long period, many hillslope processes – especially mass movements – can be treated as spatially diffuse. For example, within a brief period landslides can be seen as singularities, concentrated in one particular place. Through time, however, as the sites of individual failures shift in space reflecting the availability of susceptible material, every part of the hillslope may be subjected to this process. It is important to note, however, that diffusivity in this case applies only in the sense of spatial dsitribution. The characteristic forms that are associated with the dissipation of energy through diffusive geomorphic processes will not necessarily occur. Indeed, hillslope form will generally be linked to the dominant geomorphic process, whether this is diffusive or concentrated.

At regional scales, the suite of processes that dominate will be determined by climatic and tectonic boundary conditions (see CLIMATIC GEOMORPHOLOGY, TECTONIC GEOMORPHOLOGY). Slope angle has an important control on the manifestation of gravity. Both slope angle and other MORPHOMETRIC PROPERTIES are influenced by tectonic phenomena and lithology, in addition to the action of hillslope processes themselves. Because of the role played by water in many geomorphic processes, climate is especially important, and both the amount and variability of rainfall will influence the type of processes that occur. Soils and vegetation, as products of climatic and geological phenomena, are important, particularly in their effect on slope hydrology. Mass movements dominate on steeper slopes. Aquatic processes tend to dominate in arid to subhumid climates with gentler slopes. Despite the greater availability of water in more humid environments, initiation of aquatic processes is often precluded by protective vegetation and deeper soils with greater capacity for infiltration and subsurface runoff.

Both vegetation and soil are especially susceptible to anthropogenic influence. Although anthropogenic, in many areas tillage is recognized as an important geomorphic process in its own right (see, for example, Govers *et al.* 1994). It is diffuse, with low magnitude, and has a pronounced effect on the form of hillslopes where it occurs. Because it is mechanically initiated, tillage is neither a mass movement as defined here nor an aquatic process. A mass of material is physically displaced with each application of the plough, and in this respect tillage might be considered to be analogous with a diffuse mass movement such as soil creep. Importantly, however, the mechanical disruption and displacement of soil material can also be seen as equivalent to physical weathering, providing easily erodible material for the operation of aquatic processes.

References

Dalrymple, J.B., Blong, R.J. and Conacher, A.J. (1969) An hypothetical nine unit landsurface model, *Zeitschrift für Geomorphologie Supplementband* 12, 61–76.
Govers, G., Vandaele, K., Desmet, P.J.J., Poesen, J.W.A. and Bunte, K. (1994) The role of tillage in soil redistribution on hillslopes, *European Journal of Soil Science* 45, 469–478.

Further reading

Abrahams, A.D. (ed.) (1986) *Hillslope Processes*, Boston: Allen and Unwin.
Carson, M.A. and Kirkby, M.J. (1972) *Hillslope Form and Process*, London: Cambridge University Press.
Crozier, M.J. (1989) *Landslides: Causes, Consequences and Environment*, London: Routledge.
Dunne, T. and Leopold, L.B. (1978) *Water in Environmental Planning*, San Francisco: W.H. Freeman.
Gilbert, G.K. (1909) The convexity of hillslopes, *Journal of Geology* 17, 344–350.
Selby, M.J. (1982) *Hillslope Materials and Processes*, Oxford: Oxford University Press.
Varnes, D.J. (1978) Slope movement types and processes, in R.L. Schuster and R.J. Krizek (eds) *Landslides: Analysis and Control*, Special Report 176, 11–33, Washington, DC: Transportation Research Board, National Research Council.

SEE ALSO: freeze–thaw cycle; landslide; solifluction; threshold, geomorphic; unloading

NICK PRESTON

HOGBACK

A sharp, crested ridge of hard rock, with steeply dipping (> 20°) strata and steep near-symmetrical slopes. Hogbacks form as a result of slow differential erosion over time of alternating hard and soft strata. The soft rock is preferentially eroded, leaving steeply angled, slowly eroding resistant rock in place. The term is therefore derived from the resultant feature resembling a hog's back when viewed in planform. Examples of such features include Hogback Ridge, North Dakota, USA, Mount Rundle, Canadian Rockies and Gaishörndl, Austria.

STEVE WARD

HOLOCENE GEOMORPHOLOGY

The Holocene – or 'wholly recent' – epoch is the youngest phase of Earth history, which began with the end of the last large-scale glaciation on northern hemisphere continents other than Greenland. For this reason it is sometimes also known as the post-glacial period. In reality, however, the Holocene is one of many interglacials which have punctuated the late Cenozoic Ice Age. The Holocene is conventionally defined as beginning 10,000 radiocarbon (^{14}C) years ago, which is equivalent to about 11,500 calendar years. The term 'Holocene' was introduced by Gervais in 1869 and was accepted as part of valid geological nomenclature by the International Geological Congress in 1885. The International Union for Quaternary Research (INQUA) has a Commission devoted to the study of the Holocene, and several IGCP projects have been based around environmental changes during the Holocene. Since 1991 there has also existed a journal dedicated exclusively to Holocene research (J. Matthews, ed., *The Holocene*). A potted history of the Holocene can be found in Roberts (1998).

During the Holocene, the Earth's climates and landscapes took on their modern natural form. Geomorphological change was especially rapid during the first few millennia, with DEGLACIATION of the ICE SHEETS remaining over Scandinavia and Canada, and SEA LEVELS rising to within a few metres of their modern elevations in most parts of the world. Because soil formation and vegetation development lagged behind the often rapid shifts in climate, many landscapes – both temperate and tropical – experienced a phase of temporary geomorphic instability during this deglacial climatic transition (Thomas and Thorp 1995; Edwards and Whittington 2001). In river valleys, there were major changes in the discharge of both water and sediment, and many streams and rivers which experienced increased discharges at the end of the last glacial period are now underfit (see UNDERFIT STREAM). As a consequence of these changes, rates of denudation and sediment flux were frequently above the long-term geological norm at the start of the Holocene (Figure 85). A range of different sedimentary 'archives', including ALLUVIAL FAN, FLOODPLAIN and deltaic/estuarine deposits, can be used to establish long-term changes in rates of sediment accumulation and hence upstream soil erosion. On the other hand, river valleys are not closed systems, and quantitative SEDIMENT BUDGET calculations are more easily achieved by exploiting lake sequences. LAKES act as receptacles for materials eroded from their catchments, and dated cores can be used to calculate volumes of sediment deposited per unit time (Dearing 1994).

In most coastal regions, recognizably modern shoreline configurations were achieved around

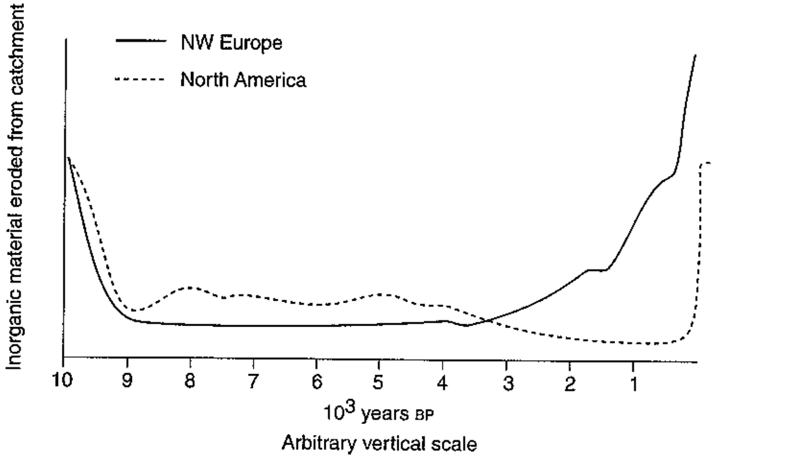


Figure 85 Generalized records of Holocene erosion based on sediment influx into lake basins (various sources, partly based on Dearing 1994)

7,000 cal. yr BP, with the main exceptions being in some high latitude regions such as Hudson Bay, where glacio-isostatic (see GLACIAL ISOSTASY) uplift has led to a continuing fall in sea levels during the Holocene. Elsewhere, rising sea levels during the early Holocene led to river valleys being drowned, with the end of this TRANSGRESSION representing the time of maximum marine incursion inland. Since then, stabilized sea levels and fluvially derived sediment discharge have led to a reversal in this trend, with the land pushing seawards at the mouths of major rivers such as the Rhône. This process has left many ancient harbour cities, such as Ephesus, Miletus and Troy in western Turkey, now stranded several km inland from the coast (Figure 86).

Various attempts have been made to subdivide the Holocene, usually on the basis of inferred climatic changes. Blytt and Sernander, for instance, proposed a scheme of alternating cool-set and warm-dry phases based on shifts in peat stratigraphy in northern Europe. In many temperate regions there is evidence of a 'thermal optimum' during the early-to-mid part of the Holocene. However, the clearest climatically induced environmental changes within the Holocene took place in the tropics and subtropics. One of the most important sources of palaeoclimatic (see PALAEOCLIMATE) and palaeohydrological (see PALAEOHYDROLOGY) information in low- and mid-latitude regions derives from non-outlet lakes, which can act like giant rain gauges. In East Africa, for example, lake levels were markedly higher and their waters markedly less saline between ~10,000 and ~6,000 cal. yr BP (Gasse 2000). On the other side, aeolian activity (see AEOLIAN PROCESSES) in regions such as the Saharan, Arabian and Thar deserts was greatly reduced and many sand dunefields (see SAND SEA AND DUNEFIELD) were inactive at this time. Rainfall in these regions increased by between 150 and 400 mm pa as tropical convectional rains moved further northwards, linked to a general strengthening of the African and Asian monsoonal system during the early Holocene.

By contrast with these climatically induced environmental changes, human impact has become an increasingly important agency in the creation and modification of landscapes during the later part of the Holocene. A critical point came when *Homo sapiens* turned to farming as the basis for human subsistence. The adoption and intensification of agriculture during the Holocene has led to widespread conversion of natural woodland or grassland into farming land, and this in turn has caused land degradation

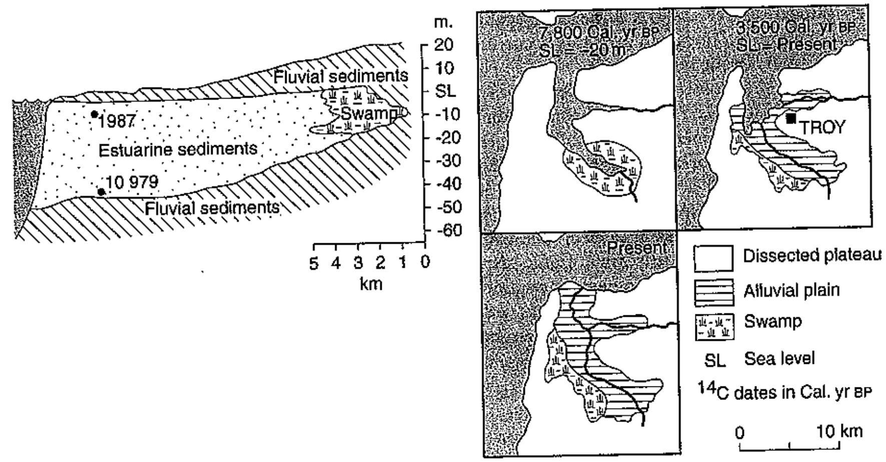


Figure 86 Geomorphological reconstructions in the vicinity of Troy, north-west Turkey, during the Holocene (based on Kraft *et al.* 1980)

through accelerated soil erosion and salinization. As a consequence, a much larger proportion of fluvial suspended sediment (see SUSPENDED LOAD) now originates from topsoil compared to bedrock sources than was previously the case. The same lake-sediment records that show high rates of flux during the major Pleistocene–Holocene climatic transition, typically show increases during the late Holocene associated with increasing human impact and land-use conversion (Figure 85). On the other hand, the dates for the onset of anthropogenically increased erosion rates vary from region to region, occurring earlier in Europe and in South and East Asia, but later in New World continents such as Australia and the Americas. At Frains Lake in the American Midwest, for example, soil erosion rates increased by two orders of magnitude to over 5 t/ha/yr^{-1} during the decade following the arrival of European settlement in 1830, but then stabilized at around 0.5–1 t/ha/yr^{-1} as forest clearance gave way to cropland (Davis 1976).

Fluctuations in climate have been superimposed upon the increasing human impact on geomorphic systems during the late Holocene. Particularly notable among these was the so-called 'Little Ice Age' from AD ~1400 to ~1850, when temperatures in Europe and the North Atlantic fell sufficiently for GLACIERS to advance down-valley in the Alps and other mountains

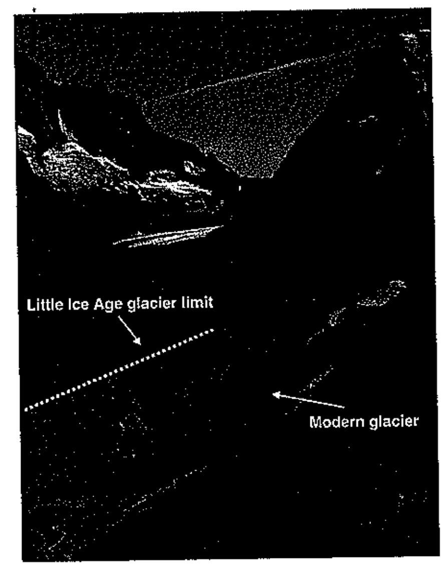


Plate 61 Modern and Little Ice Age limits of the Lower Arolla glacier, Switzerland

(Plate 61). This climatic deterioration brought a higher risk of geomorphological hazards including LANDSLIDES, avalanches, glacier outbursts and other floods (Grove 2003), with comparable periods of extreme floods and droughts occurring in dryland parts of the world.

It is during the Holocene that modern boundary conditions for the Earth system have come into existence. Consequently the Holocene represents a key baseline for assessing human impact on Earth surface and atmospheric processes, in our attempts to tease apart the relative roles of natural and human agencies in rates of landscape change. It also provides the time frame over which long-term magnitude–frequency relationships (see MANGNITUDE–FREQUENCY CONCEPT) can be assessed and the return period for extreme events, such as floods, can be calculated (Benito *et al.* 1998; Knox 2000).

References

Benito, G., Baker, V.R. and Gregory, K.J. (eds) (1998) *Palaeohydrology and Environmental Change*, Chichester: Wiley.

Davis, M.B. (1976) Erosion rates and land use history in southern Michigan, *Environmental Conservation* 3, 139–148.

Dearing, J. (1994) Reconstructing the history of soil erosion, in N. Roberts (ed.) *The Changing Global Environment*, 242–261, Oxford: Blackwell.

Edwards, K.J. and Whittington, G. (2001) Lake sediments, erosion and landscape change during the Holocene in Britain and Ireland, *Catena* 42, 143–173.

Gasse, F. (2000) Hydrological change in the African tropics since the Last Glacial Maximum, *Quaternary Science Reviews* 19, 189–212.

Grove, J.M. (2003) *The Little Ice Ages: Ancient and Modern*, Routledge: London.

Knox, J.C. (2000) Sensitivity of modern and Holocene floods to climate change, *Quaternary Science Reviews* 19, 439–458.

Kraft, J.C., Kayan, I. and Erol, O. (1980) Geomorphic reconstructions in the environs of ancient Troy, *Science* 209, 776–782.

Roberts, N. (1998) *The Holocene. An Environmental History*, 2nd edition, Blackwell: Oxford.

Thomas, M.F. and Thorp, M.B. (1995) Geomorphic response to rapid climatic and hydrologic change during the Late Pleistocene and Early Holocene in the humid and sub-humid-tropics, *Quaternary Science Reviews* 14, 101–124.

NEIL ROBERTS

HONEYCOMB WEATHERING

Honeycomb weathering is a type of CAVERNOUS WEATHERING. The terms honeycomb, stone lattice, stone lace and alveolar weathering have been used as synonyms. Honeycombs are associated in

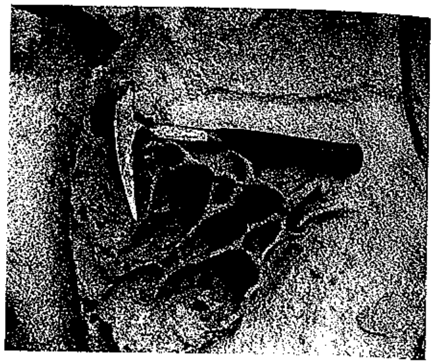


Plate 62 A small cluster of alveoles or honeycomb weathering forms developed on granite gneiss in a salty and foggy coastal environment on the south coast of Namibia, near Luderitz

particular with arid and coastal environments, though they are also a feature of some building stones in urban environments. They may also occur on Mars (Rodriquez-Navarro 1998). Many honeycombs are seemingly caused by SALT WEATHERING (Mustoe 1982; Rodriguez-Navarro *et al.* 1999). They are composed of pits, commonly some centimetres deep, that are developed so close together as to be separated by a narrow wall only millimetres thick. They are known from a wide range of rock types, including sandstones, limestone, schists, gneiss, greywacke, arkose and metavolcanics. In favourable environments they can form in a matter of decades (Mottershead 1994).

References

Mottershead, D.N. (1994). Spatial variations in intensity of alveolar weathering of a dated sandstone structure in a coastal environment, Weston-Super-Mare, UK, in D.A. Robinson and R.B.G. Williams (eds) *Rock Weathering and Landforms Evolution*, 151–174, Chichester: Wiley.

Mustoe, G.E. (1982). The origin of honeycomb weathering, *Geological Society of America Bulletin* 93, 108–115.

Rodriguez-Navarro, C. (1998) Evidence of honeycomb weathering on Mars, *Geophysical Research Letters* 25, 3,249–3,252.

Rodriquez-Navarro, C., Doehne, E. and Sebastian, E. (1999) Origins of honeycomb weathering; the role of salts and wind, *Geological Society of America Bulletin* 111, 1,250–1,255.

A.S. GOUDIE

HOODOO

A common North American term for a pillar of eroded rock which is capped by a resistant rock layer, protecting a column of more erodable material beneath. They are effectively remnants of steep slopes as they are being eroded back, forming as the less resistant material is eroded away by water. The overlying hard caprock (generally a boulder or cobble) maintains the form's vertical integrity. Hoodoos are common in BADLAND morphology, and are typically formed in sedimentary rock (e.g. Bryce Canyon, Utah, USA) although examples exist in volcanoclastics (e.g. San Juan Mountains of Colorado, USA) and unconsolidated glacifluvial materials (e.g. Norway).

SEE ALSO: demoiselle

STEVE WARD

HORST

A relatively upraised fault block bounded by sharply defined and sometimes parallel reverse faults, though more commonly conjugate normal faults and opposing dips. The formation of the horst can be due to both the rifting and compressive movement of these marginal normal faults. Horsts are generally elongate ridge-like structures, with a plateau form on the uplifted horst block surface. Horst is a German term, and means retreat. Converse to horsts are graben (singular and plural). These are relatively low-standing fault blocks once again bounded by opposing normal faults, and occurring between zones of extension or compression. Half-grabens are grabens that are bounded by a normal fault on one side only. Areas of alternating uplifted and down-dropped fault blocks are thus referred to as horst and graben structures, and are associated with RIFT VALLEY AND RIFTING. In these regions, horsts are often the predominant sediment source into the down-dropped graben and any basins within. Examples of horst structures include the Black Forest and Harz Mountains in Germany, and the Vosges of eastern France. A famous graben structure is the Rhine Graben in Germany.

Further reading

Jaroszewski, W. and Kirk, W. L. (1984) *Fault and Fold Tectonics*, Chichester: Ellis Horwood.

SEE ALSO: fault and fault scarp

STEVE WARD

HORTON'S LAWS

In 1945, in one of the most significant twentieth-century contributions to geomorphology, the American engineer, Robert E. Horton endeavoured to express both the hierarchical arrangement and density of drainage networks in quantitative terms. In this he was explicitly following what he termed the 'ocular observation' (Horton 1945: 280) of the Scottish mathematician John Playfair (1802: 102).

> Every valley appears to consist of a main trunk, fed from a variety of branches, each running in a valley proportioned to its size, and all of them together forming a system of vallies, communicating with one another, and having such a nice adjustment of their declivities that none of them join the principal valley either on too high or too low a level.

Horton proceeded by introducing, first, the concept of STREAM ORDERING (1945: 281), in the following fashion:

> (U)nbranched fingertip tributaries are always designated as of order 1, tributaries or streams of the 2nd order receive branches or tributaries of the 1st order, but these only; a 3rd order stream must receive one or more tributaries of the 2nd order but may also receive 1st order tributaries. A 4th order stream receives branches of the 3rd and usually of lower orders, and so on. Using this system, the order of the main stream is the highest.

Somewhat unfortunately, Horton then developed an approach whereby the 'parent stream' had to be identified from source to mouth (Horton 1945: figure 7), something which is tricky and undoubtedly subjective (Figure 87a). A.N. Strahler (1957) proposed a modification of Horton's scheme (Figure 87b) which is less subjective and has been almost universally applied since 1957.

From the stream-ordering system, Horton proceeded to calculate two indices which he found to evince such regularity that they have become known as Horton's Laws (they are equally apparent whichever ordering scheme is used). The first observation was that the number of streams of different orders tended to follow an inverse geometric sequence. The ratio between each order being termed r_b, the bifurcation ratio. The second was that the average length of streams tended to increase as order rose. One set of Horton's data (and the value of r_b and r_l) are given in Table 25. If the data for stream numbers and lengths are plotted on semi-logarithmic paper, straight lines can be drawn showing roughly constant ratios throughout any one basin.

One further law – of stream slopes – was described by Horton (1945: 295), showing a geometric decrease in channel slope with increasing stream order; and yet two further laws (of basin area, which increases regularly with order) and the constant of channel maintenance were introduced by Schumm (1956). Horton's three basic 'laws' were set out (1945: 84 and figure 6) as:

Law of stream numbers: $N_o = r_b^{(s-o)}$
Law of stream lengths: $l_s = l_1 r_l^{0-1}$
Law of stream slopes: $s_c = s_1 \sqrt{r_s}^{(0-1)}$

Where: N_o is the number of streams in the drainage basin
r_b is the bifurcation ratio
s is the order of the main stream
o is the order of a given class of tributaries
l_s is the length of the main stream
l_1 is the average length of first-order streams
r_l is the length ratio
s_c is the slope of the main stream
s_1 is the average slope of first-order streams
r_s is the slope ratio

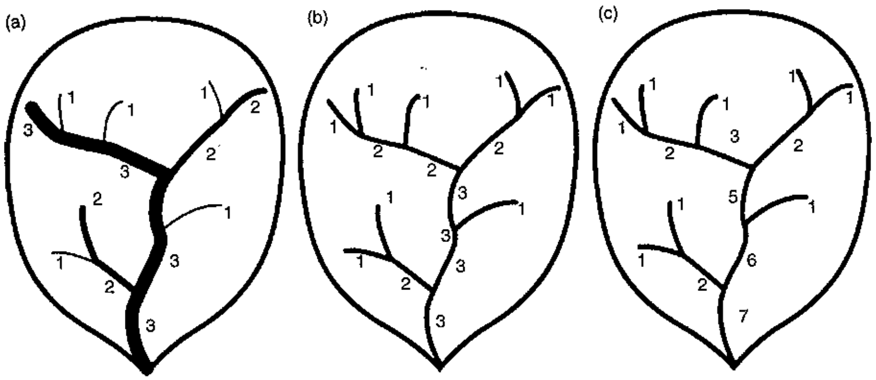


Figure 87 A drainage network with channels ordered: (a) according to Horton (1945); (b) according to Strahler (1957); (c) according to Shreve (1966)

Table 25 Drainage net, upper Hiwassee River

Order	No. of streams	Bifurcation ratio r_b	Average length, miles	Ratio of length, r_l
1	146		0.49	
		4.6		2.6
2	32		1.28	
		3.6		2.85
3	9		3.65	
		3.6		3.37
4	2		12.30	

Source: from Horton (1945: figure 7)

Schumm's additions may be stated in his words as: 'the mean drainage-basin areas of streams of each order tend to approximate closely a direct geometric series in which the first term is the mean area of the first-order basins' (1956: 606); and 'the relationship between mean drainage-basin areas of each order and mean channel lengths of each order is a linear function whose slope. . . is equivalent to the area in square feet necessary. . . for the maintenance of 1 foot. . . of channel' (1956: 607).

As Rodríguez-Iturbe and Rinaldo (1997: 6) make clear, the Hortonian laws, over the years, have been considered variously as demonstrating: that basins show a regular, evolutionary process; that basins demonstrate the development of purely topologically random networks; and that they say nothing whatsoever except that virtually all networks show these relationships. This last point is linked to the intuitively puzzling fact that Horton's Laws emerge whether basins are described by Horton's or by Strahler's methods.

However, there is a more fundamental difficulty with Horton's analysis as described in 1945: the categories of stream channels plotted on the x axis of the semi-logarithmic plots are, of course, *ordinal* numbers. As Horton's definition, quoted above and Figures 87a and b make plain, a 'second-order' basin may contain any number of 'first-order' channels between 2 and (effectively) ∞. Whilst it is quite proper to identify basins of different orders and to express numbers and averages in each class of variables and even ratios of values between classes, it is not proper to conduct the mathematical operations of multiplication or division using ordinal numbers and, in consequence, the 'regressions' shown on Horton's 1945 plots are simply spurious.

However, Horton introduced two other quantitative measures: DRAINAGE DENSITY (D_d) – the length of stream channels per unit area – and stream frequency (F_s), or the number of streams per unit area. Melton (1958) demonstrated that these two terms could be linked by a constant relationship:

$$F_s = 0.694\, D_d^2$$

This is now known as Melton's Law (Rodríguez-Iturbe and Rinaldo 1997: 8). These relationships described the degree of dissection of a landscape (see Kennedy 1978).

But the importance of the emphasis on stream frequency, involving actual counting of entities, was crucially developed by R. Shreve (1966) as the concept of basin magnitude, which is simply the number of first-order streams (or exterior links). Figure 87c shows the nature of this, true, ordering system, where magnitude is a ratio number that is susceptible of full mathematical manipulation.

Despite the difficulties with the Horton/Strahler ordering system, it is far easier to establish a sample of (say) fourth order basins than of 10 magnitude ones and a huge volume of research, especially although not exclusively from the 1950s and 1960s, focused on Hortonian relationships in generally fourth- or fifth-order basins. Church and Mark (1980) showed how this focus had tended to produce an apparent scale-dependence in the relationship between basin area and drainage density, which is actually isometric.

Nevertheless, the thrust of recent work on the FRACTAL nature of drainage basins, notably that summarized by Rodríguez-Iturbe and Rinaldo (1997), has been to substantiate the universality of the two fundamental Hortonian Laws: for example, their discussion of Optimal Channel Networks (OCNS) is shown to support both the Bifurcation Law and the Length ratio (1997: 278–279). Indeed, there is a major discussion of Hortonian networks which 'can be interpreted as signs of the fractal structure of the underlying network' (1997: 498). It is worth stressing, however, that what has endured have been the geometrically regular ratios of stream numbers and lengths between adjacent stream orders, rather than the spurious 'regression' plots which were such a feature of mid-twentieth century investigations of Horton's Laws.

References

Church, M.A. and Mark, D.M. (1980) On size and scale in geomorphology, *Progress in Physical Geography* 4, 342–390.
Horton, R.E. (1945) Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, *Geological Society of America Bulletin* 56, 275–370.
Kennedy, B.A. (1978) After Horton, *Earth Surface Processes and landforms* 3, 219–232.
Melton, M.A. (1958) Correlation structure of morphometric properties of drainage systems and their controlling agents, *Journal of Geology* 66, 442–460.
Playfair, J. (1802) *Illustrations of the Huttonian Theory of the Earth*, London: Cadell and Davies. Reprinted in facsimile, G.W. White (ed.) (1964), New York: Dover.

Rodríquez-Iturbe, I. and Rinaldo, A. (1997) *Fractal River Basins: Chance and Self-organization*, Cambridge: Cambridge University Press.
Schumm, S.A. (1956) Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey, *Geological Society of America Bulletin* 67, 597–646.
Shreve, R.L. (1966) Statistical law of stream numbers, *Journal of Geology* 74, 17–37.
Strahler, A.N. (1957) Quantitative analysis of watershed geomorphology, *EOS Transactions American Geophysical Union* 38, 912–920.

Further reading

Knighton, D. (1998) *Fluvial Forms and Processes*, 2nd edition, London: Arnold.
Schumm, S.A. (ed.) (1977) *Drainage Basin Morphology*, Stroudsburg, PA: Dowden, Hutchinson and Ross.

SEE ALSO: drainage density; fractal; laws, geomorphological; stream ordering

BARBARA A. KENNEDY

HUMMOCK

Small mounds of low relief, which cover the ground surface, are common where fine-grained soils overlie PERMAFROST. Most hummocks are circular, and 1 to 2 m in diameter. They are domed, with a vertical relief of up to 25 cm, but usually less than 15 cm. The ACTIVE LAYER is thickest beneath the hummock centres, and thinnest at the circumference. The base of the active layer is bowl-shaped. Segregated ice lenses, subparallel to the base of the active layer, are characteristically abundant directly beneath hummocks, and this ice-rich zone is commonly also rich in organic material. Hummocks are generally stable features, which may persist for thousands of years.

At the surface, organic material accumulates around the edges of hummocks, but the centres may be bare of vegetation (mud hummocks) or covered by peat or vascular plants (earth hummocks). The soil in hummocks is frost-susceptible, and may contain little sand. Where the clay content is low, the soil may liquefy in response to small changes in moisture content of stress, and be extruded at the ground surface. Such mudboils may occur in fields of hummocks, but in general the clay content is of the order of 40 to 50 per cent, and sufficient to prevent liquefaction.

The hummock form is maintained by soil circulation within each feature, driven by moisture redistribution during freezing and thawing (Mackay 1980). The soil circulation proceeds by upwards movement in the middle of hummocks, spreads to the circumference near the surface, and slides downwards at the edges, near the base of the active layer. The upward movement is driven by convection, due to the contrast between mud of relatively low density, and enclosing sediment, where the mud is supersaturated by melting of ice lenses. The movement at the base of the active layer is associated with heave towards the hummock centre during upfreezing at the base of the active layer, and settlement down the bowl-shaped frost table as ice lenses thaw. At the surface, soil is driven outward by heave and subsidence during freezing and thawing over an inclined plane. These three processes are constrained by the requirement for conservation of mass. Evidence for the circulation has been provided by movement of markers at the ground surface, and by involutions in the soil stratigraphy when viewed in cross section. The importance of the bowl-shaped frost table on hummock form is demonstrated by the disappearance of hummocks in the years following forest fires, when the active layer deepens, and their reappearance with subsequent vegetation regeneration as the active layer thins.

In the boreal forest, trees on hummocks are tilted, and are commonly located near the edges of hummocks. Tilting of the trees is associated with development of the ice-rich zone at the base of the active layer and accompanying heave of the hummocks. Hummocks are the complementary feature for fine-grained soil of sorted circles in coarser materials.

Reference

Mackay, J.R. (1980) The origin of hummocks, western Arctic coast, Canada, *Canadian Journal of Earth Sciences* 17, 996–1,006.

C.R. BURN

HYDRATION

Hydration is the uptake of the entire water molecule by a mineral. For example, calcium sulphate (anhydrite $CaSO_4$) is hydrated to gypsum $CaSO_4 \cdot H_2O$. This results in the mineral swelling. In a confined space, hydration pressures can be up to 100 Mpa, weakening the rock. In cold climates, White (1976) felt that much freeze–thaw weathering could actually be hydration shattering, with the forces of hydration as high as 2,000 kg cm^{-3}.

Widely occurring is the conversion of iron oxide (Haematite Fe_2O_3) into iron hydroxides variously cited as being in a poorly defined crystal form as $Fe(OH)_3$, as goethite $2FeOOH$ or as limonite $(2Fe_2O_3 \cdot 3H_2O)$. The formation of these iron hydroxides involves considerable volume increases.

Alumino-silicate minerals can become subject to hydration through the formation of hydrated aluminium oxide. HYDROLYSIS can be seen as more important than hydration because it is the products of hydrolysis which are hydrated. For the formation of the hydrated aluminium oxide from microcline, a potassium-containing feldspar:

$$KAlSi_3O_8 + H_2O \rightarrow HAlSi_3O_8 + K^+ + OH^- \quad \text{(hydrolysis)}$$

Microcline

$$2HAlSi_3O_8 + 11H_2O \rightarrow Al_2O_3 + 6H_4SiO_4 \quad \text{(hydrolysis)}$$

$$Al_2O_3 + 3H_2O \rightarrow Al_2O_3 \cdot 3H_2O \quad \text{(hydration)}$$

However, it should be stressed that it is the hydration which facilitates the physical disintegration through volume expansion, weakening the mineral structure.

In addition to this formation of new hydrated minerals, more complex layered minerals can take up water between their layers and this can also be referred to as hydration. The plate-like minerals, such as mica, can be subject to expansion and physical disintegration when water penetrates between the plates. Water can be incorporated into a clay crystal lattice and, especially in the open lattice of montmorillonite clays, involving increases in volume of around 0.5 cm^3 g^{-1}.

Reference

White, S.E. (1976) Is frost action really only hydration shattering? *Arctic and Alpine Research* 8, 1–6.

STEVE TRUDGILL

HYDRAULIC GEOMETRY

Hydraulic geometry of a river is the quantitative (mathematical and graphical) description of the channel cross section size and shape, fluid-flow properties and sediment-transport characteristics, in relation to the discharge being conducted by the channel. As such, every river channel, with rigid or deformable boundaries, has a hydraulic geometry. It is a descriptive device, derived from the empirical relations of regime 'theory' developed to aid canal design in India early last century (Lacey 1929). These ideas were first introduced into geomorphology by Leopold and Maddock (1953) who proposed the term 'hydraulic geometry' for this descriptor of the morphodynamics of alluvial channels.

The general equations of hydraulic geometry proposed by Leopold and Maddock (1953) necessarily are selective, reflecting relations among variables that were routinely measured or easily derived from such measurements made at US gauging stations:

$$\begin{aligned} w &= aQ^b \\ d &= cQ^f \\ v &= kQ^m \\ s &= gQ^z \\ n &= tQ^y \\ ff &= hQ^p \\ Q_{susp} &= rQ^j \end{aligned}$$

where w, d, v, s, n, ff and Q_{susp} are respectively width, mean depth, mean velocity, water-surface slope, flow resistance (Manning's n or D'Arcy Weisbach ff) and suspended-sediment load. An important missing element of this seven-variable set is bedload transport but these measured data are rarely available.

Implicit in the specification of these equations of hydraulic geometry are the following notions:

1. Discharge, Q, is the dominant independent variable in the hydraulic geometry;
2. The relations between the independent and dependent variables can be described as simple power functions;
3. As power functions, the logarithm of the dependent variables plot against the logarithm of discharge as a straight-line graph (that is, there is a linear relationship between the order-of-magnitude increases in the pairs of variables);
4. The existence of these orderly hydraulic-geometry relations implies an underlying set of processes reflecting the operation of equilibrium in the morphodynamic system;
5. Because continuity must be satisfied in fluid flow it follows from the rules of algebra that:

$$Q = wdv = (aQ^b)\,(cQ^f)\,(kQ^m)$$
and that $ack = 1$ and $b + f + m = 1$

The adjustment of channel morphology and hydraulics in relation to changes in discharge has been considered in two quite different contexts: at-a-station hydraulic geometry and downstream hydraulic geometry.

At-a-station hydraulic geometry

The at-a-station hydraulic geometry describes how the channel geometry and hydraulics of flow change as discharge increases at an individual channel cross section of a river. Thus it describes the way in which the flow fills the often effectively rigid channel boundary as discharge changes over time. An example of this type of hydraulic geometry is shown in Figure 88 for the Fraser River in western Canada. At-a-station hydraulic geometry is only defined for discharges up to the channel-filling (or bankfull) stage.

Downstream hydraulic geometry

Discharge in a river also increases as tributaries join the main stem in the downstream direction and add flow to the fluvial system. The downstream hydraulic geometry describes how this spatially increasing discharge enlarges and shapes the channel and alters the properties of the streamflow. In order to allow for comparisons between channel sections these changes are referred to as a discharge of constant return period or to a consistent relative stage. The most common reference discharge is bankfull discharge, which is often taken to be the channel-forming discharge. An example of this type of hydraulic geometry is shown in Figure 89 for Oldman River in western Canada.

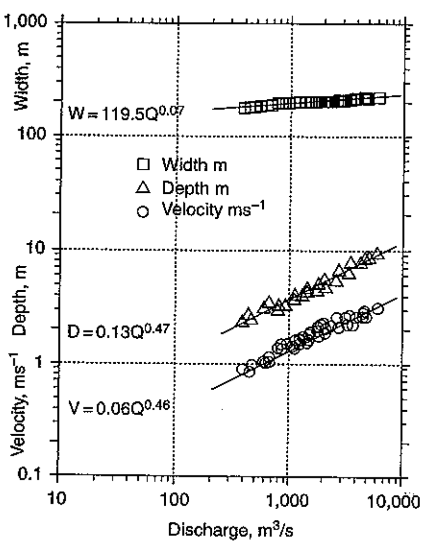


Figure 88 Hydraulic geometry for the Fraser River, Canada

Theoretical context and interpretation of hydraulic geometry

Conventional hydraulic geometry describes a partial picture of adjustments in the fluvial system but contains little information about the controls on such adjustments. When responding to changes in discharge, an alluvial stream must satisfy at least three sets of physical relations: continuity, flow resistance, and sediment transport. The first relation is definitional but the other two relations are only understood in a qualitative sense. For this reason, and also because channels are free to adjust in ways other than by changes in width, depth and mean velocity (Hey 1988), the hydraulic geometry of alluvial channels generally is regarded as being indeterminate. Nevertheless, these physical relations do inform our interpretation of the hydraulic geometry at a qualitative level of analysis.

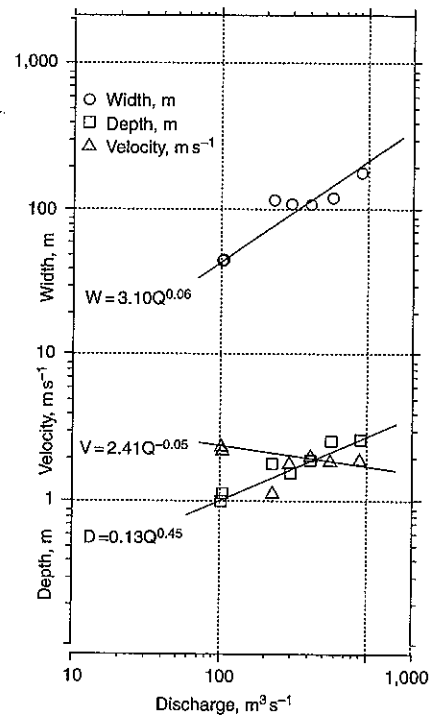


Figure 89 Hydraulic geometry for the Oldman River, Canada

The primary difference between the two hydraulic geometries is that, unlike the downstream case, for the at-a-station case the boundary materials and the water-surface slope essentially remain constant as discharge changes. The principal control on the hydraulic geometry is the shape of the channel cross section formed at high discharges. Channel form in turn largely is determined by the strength of the boundary materials. If the boundary materials are cohesive and strong (mud-dominated rivers, for example), banks can develop which are steep and high. In such cases, as discharge increases, channel width will change much more slowly than flow depth and velocity. The rate of increase in flow velocity depends on the changing relative roughness of the channel. Typically, but not always, flow resistance declines as the effects of roughness elements are drowned by the increasing discharge. Thus the exponents on the velocity/discharge relation are relatively high. A different geomorphic response can be expected in the case of channels with weak non-cohesive banks (such as sand-dominated channels). In this case bank height is limited by material strength and the capacity for mean depth changes to accommodate increases in discharge is small. Change in velocity is also highly constrained by the conservative adjustment in flow depth and as a result channel width changes greatly to accommodate the discharge increases in such channels. Table 26 shows typical exponent values in the at-a-station hydraulic geometry equations for a variety of channel types.

In the case of downstream hydraulic geometry, adjustments of river channels to increases in bankfull discharge also reflect spatial changes in water-surface slope and the size of boundary-materials along the channel. A major control in this case is the balance that is struck between the impelling and resisting forces acting in the downstream direction. Although downstream increases in discharge and declining size of boundary material tend to work together to accelerate the flow, the forces for this change are almost equally countered by the decline in boundary shear stress related to declining water-surface slope. As a result, downstream hydraulic geometry is characterized by constant or declining mean velocity downstream and by the necessity for changes in discharge to be almost fully accommodated by adjustments in channel width and mean flow-depth alone. As before, the apportioning of the discharge change between width and velocity depends largely on boundary material strength. In a mud-dominated channel which can support steep and high channel banks the exponent on depth will be high and that on width low. In contrast, a sand-bed channel with weak banks will accommodate most of the increased bankfull

Table 26 Selected values of exponents in the equations of hydraulic geometry of river channels

Channel locality and type	At-a-station values			Downstream values		
	b	f	m	b	f	m
Mid-west USA (Leopold and Maddock 1953)	0.26	0.40	0.34	0.50	0.40	0.10
Mid-west USA (Carlston 1969)				0.46	0.38	0.16
Ephemeral streams, semi-arid USA (Leopold and Miller 1956)	0.25	0.41	0.33	0.50	0.30	0.20
Upper Salmon River, Idaho (Emmett 1975)				0.54	0.34	0.12
R. Bollin Dean, coarse-bed cohesive banks (Knighton 1974)	0.12	0.40	0.48	0.46	0.16	0.38
British gravel-bed rivers (Charleton *et al.* 1978)				0.45	0.40	0.15
Columbia River, Canada, canal-like anastomosed sand channels, cohesive banks (Tabata 2002)	0.10	0.66	0.24			

Notes: $W = aQ^b$, $D = cQ^f$, $V = kQ^m$

discharge downstream by widening the channel and the exponent for depth will be correspondingly low. Table 26 shows typical exponent values in the downstream hydraulic geometry equations for a variety of channel types.

Comprehensive reviews of all aspects of the hydraulic geometry of natural stream channels are available in fluvial geomorphology texts such as those by Knighton (1998), Richards (1982) and Leopold *et al.* (1964).

Limitations of hydraulic geometry

The power of conventional hydraulic geometry is its facility to generalize the process of channel adjustment in terms of simple functions so that the morphology of channels can be compared readily, one with another. This facility is also its primary limitation: it ascribes to the channel adjustment process the simple behaviour of simple functions which in detailed reality, however, may be very complex. Given the typical statistical noise in measurements of hydraulics and morphology of natural channels, power functions provide a simple and robust model for the hydraulic geometry but there is no independent theoretical justification for their use. The power-function model is merely a convenient approximation of reality.

Geomorphologists who recognize the limitations of power functions and have suggested hydraulic geometries based on alternative linearizing models (such as the log-quadratic model) include Richards (1973), Knighton (1975) and Ferguson (1986). Still others have taken multivariate statistical approaches to characterizing hydraulic geometry (Bates 1990; Rhoads 1992).

All these mathematical models of hydraulic geometry imply, however, that the channel adjustment process is continuous when in reality it is often markedly discontinuous. For example, many channels exhibit channel-in-channel morphology reflecting the capacity of low flows to shape the basal boundary (long duration offsetting low flow magnitude). Others exhibit within-channel benches reflecting the unequal effectiveness of higher discharge ranges of the hydrologic regime in shaping the channel (Woodyer 1968). Further, although low discharges essentially flow over rigid boundaries in many channels, increases in discharge at a section eventually lead to the onset of sediment entrainment and channel scour and this fundamentally alters the morphodynamics of the channel. Indeed such scour-related discontinuities in channel adjustments to changes in discharge may be an important part of the adjustment regime but may be completely obscured by the use of hydraulic geometry (Hickin 1995).

In the case of downstream hydraulic geometry the point-source addition of tributary flow and sediment to the fluvial system is a fundamental spatial and process discontinuity that is essentially a step-function process, not the continuous adjustment approximated by power functions.

References

Bates, B.C. (1990) A statistical log piecewise linear model of at-a-station hydraulic geometry, *Water Resources Research* 26, 109–118.

Carlston, C.W. (1969) Downstream variations in the hydraulic geometry of streams: special emphasis on mean velocity, *American Journal of Science* 267, 499–510.

Charleton, F.G., Brown, P.M. and Benson, R.W. (1978) The hydraulic geometry of some gravel rivers in Britain, *Hydraulics Research Station Report*, IT 180.

Emmett, W.W. (1975) The channels and waters of the Upper Salmon River area, Idaho, *US Geological Survey Professional Paper* 870A.

Ferguson, R.I. (1986) Hydraulics and hydraulic geometry, *Progress in Physical Geography* 10, 1–31.

Hey, R.D. (1988) Mathematical models of channel morphology, in M.G. Anderson (ed.) *Modelling Geomorphological Systems*, 99–125, Chichester: Wiley.

Hickin, E.J. (1995) Hydraulic geometry and channel scour: Fraser River, B.C., Canada, in E.J. Hickin (ed.) *River Geomorphology*, 155–167, Chichester: Wiley.

Knighton, A.D. (1974) Variation in width-discharge relation and some implications for hydraulic geometry, *Geological Society of America Bulletin* 85, 1,069–1,076.

——(1975) Variations in at-a-station hydraulic geometry, *American Journal of Science* 275, 186–218.

——(1998) *Fluvial Forms and Processes: A New Perspective*, London: Arnold.

Lacey, C. (1929) Stable channels in alluvium, *Proceedings of the Institution of Civil Engineers* 229, 259–384.

Leopold, L.B. and Maddock, T. (1953) The hydraulic geometry of stream channels and some physiographic implications, *United States Geological Survey Professional Paper* 252.

Leopold, L.B. and Miller, J.P. (1956) Ephemeral streams – hydraulic factors and their relation to the drainage net, *United States Geological Survey Professional Paper* 282A.

Leopold, L.B., Wolman, M.G. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: Freeman.

Rhoads, B.L. (1992) Statistical models of fluvial systems, *Geomorphology* 5, 433–455.

Richards, K.S. (1973) Hydraulic geometry and channel roughness – a non-linear system, *American Journal of Science* 273, 877–896.

Richards, K.S. (1982) *Rivers: Form and Process in Alluvial Rivers*, London: Methuen.

Tabata, K.K. (2002) Character and conductivity of anastomosing channels, upper Columbia River, British Columbia, Canada, M.Sc. Thesis, Department of Geography, Simon Fraser University, BC, Canada.

Woodyer, K.D. (1968) Bankfull frequency in rivers, *Journal of Hydrology* 6, 114–142.

Further reading

Kellerhals, R., Neill, C.R. and Bray, D.I. (1972) Hydraulic and geomorphic characteristics of rivers in Alberta, *Research Council of Alberta, River Engineering and Surface Hydrology Report* 72–1, 16–18.

SEE ALSO: channel, alluvial; fluvial geomorphology

EDWARD J. HICKIN

HYDRO-LACCOLITH

A mound of ice formed by frost heaving of frozen underground water, resembling a laccolith in section. The term hydro-laccolith is synonymous to the terms ice laccolith and PINGO. However, they differ from pingos in that they are seasonal forms (whereas pingos are perennial), and differ from ice laccoliths in that they do not form within the active layer of permafrost ground. Hydro-laccoliths range in size between 1 and 10 m diameter, and are usually less than 2 m in height.

Further reading

French, H.M. (1996) *The Periglacial Environment*, Harlow: Longman.

SEE ALSO: periglacial geomorphology

STEVE WARD

HYDROCOMPACTION

The compaction and reduction in volume of soils and sediments that occurs when their moisture content is increased. It is also known as 'collapse compression', 'hydrocompression', 'hydroconsolidation' and 'saturation shrinkage' (Charles 1994). The process causes ground subsidence when unconsolidated sediments of low density are wetted, as for example by the application of irrigation water. It is a feature of arid and semi-arid lands where materials such as wind-blown loess or certain alluvial sediments above the water table are not normally wetted below the root zone and have high void ratios. When dry, such materials may have sufficient strength to support considerable effective stresses without compacting. However, when they are wetted, their intergranular strength is weakened because of the rearrangement of their particles. The associated subsidence may create fissures in the ground and is a process that needs to be considered during the construction of canals, pipelines, dams and irrigation schemes (Al-Harthi and Bankher 1999).

References

Al-Harthi, A.A. and Bankher, K.A. (1999) Collapsing loess-like soil in western Saudi Arabia, *Journal of Arid Environments* 41, 383–399.

Charles, J.A. (1994) Collapse compression of fills on inundation, in K.R. Saxena (ed.) *Geotechnical Engineering: Emerging Trends in Design and Practice*, 353–375, Rotterdam: Balkema.

A.S. GOUDIE

HYDROLOGICAL GEOMORPHOLOGY

Hydrological geomorphology is literally the interface between hydrology, the science of water and geomorphology, the study of landforms and their causative processes. It is particularly surface water hydrology that interacts with geomorphology although recently there has been an increasing convergence between research in geomorphology and in groundwater hydrology and hydrogeology (Brown and Bradley 1995).

When geomorphology emerged, more than a century ago, with a focus upon the morphology of landscape and the study of landforms, the attractive CYCLE OF EROSION promulgated by W.M. Davis provided a focus for many approaches to geomorphology adopted in the first half of the twentieth century. In the second half of the century, investigation of processes prevailed to a much greater extent (Gregory 2000) and to achieve this in relation to research on land areas, involving the study of flowing water, there was a need to take a much greater interest in hydrology. At first a period of familiarization saw the publication of books written by geographers (e.g. Ward 1966) and geomorphologists (e.g. Gregory and Walling 1973) but then progress was made towards the research interface between geomorphology and hydrology where the geomorphologist could make significant contributions. Hydrology was for long the science of water, with comparatively little attention given to water quality, but increasing attention given to landscape-forming processes, and to hydrological influences upon those processes, naturally led to original and innovative contributions being made by geomorphologists at the interface of hydrology with geomorphology.

Many interface investigations have accounted for the growth of hydrological geomorphology and these include at least four types. In addition to relationships between drainage basin characteristics and basin hydrological response, geomorphologists have made particular contributions in the investigation of *runoff producing areas* and the dynamic ways in which such areas contribute to the generation of stream hydrographs, including networks of subsurface pipes as well as headwater drainage systems and the modelling of their role in runoff production (Beven and Kirkby 1993). Such investigations often employed results from small experimental catchments which were also useful in relation to research on *sediment dynamics*. Geomorphological interest in the sediment area arose from the requirement for rates of erosion or denudation to relate to landscape development. In the case of suspended sediment production and transport, whereas simple rating curves relating discharge and suspended sediment concentrations had previously been used for analysis, it was demonstrated how analysis of sediment hydrographs could be employed to advance understanding and explanation of the mechanics of erosion. Later research by geomorphologists could therefore focus on the mechanics of production of such sediment in relation to the range of sediment sources and sediment producing areas. Similar contributions, made by geomorphologists to investigate and refine relationships employed to model bedload and solute transport, benefited from studies of the generation of solutes from catchment areas and of the entrainment of bedload in different channel situations.

Contributions at the level of the drainage basin have arisen first because of interest in the *temporal changes* that have occurred. As such changes cannot be based entirely on continuous hydrological records used in hydrology, other techniques are necessary to reconstruct past hydrological changes and these are used in geomorphology. The approach of palaeohydrology (Schumm 1965), which has been defined as 'the science of the waters of the earth, their composition, distribution and movement on ancient landscapes from the beginning of the first rainfall to the beginning of continuous hydrological records' (Gregory 1996), has been developed significantly so that the broad picture of past hydrological changes has been reconstructed for different parts of the world (Benito and Gregory 2003). Such reconstructions can provide potentially useful background for studies of global change and of basin management. A particularly valuable and successful approach has been the investigation and analysis of PALAEOFLOODS based upon the recognition of remnants of flood deposits, often as slackwater deposits, because this affords information on flood frequency which extends well beyond the period of instrumental record but may significantly affect the way in which flood frequency is analysed and relationships are established (Baker 2003). Partly as a result of the results obtained from investigations of temporal change, geomorphological contributions have become significant in relation to *river basin management*. Approaches are increasingly required to be integrated and should therefore include consideration of the range of human impacts throughout the basin (Downs *et al.* 1991) but in addition there is a need for sustainable approaches both at the basin level (NRC 1999) and in relation to the restoration of specific river reaches (Brookes and Shields 1996).

A number of the salient contributions made by geomorphologists have illuminated understanding of particular components of the hydrological cycle so that aspects of a hydrological geomorphology have emerged. This has, paradoxically, led to the fudging of the original definition of geomorphology. No longer focused primarily upon landforms, geomorphology is involved in contributions to hybrid fields where some of the most innovative research occurs and where multidisciplinary approaches can be optimized. Thus hydromorphology was suggested by Scheidegger (1973) as the geomorphological study of water and its effects, which includes coastal as well as fluvial hydrogeomorphology, in which there is a range of ways in which applications can be made (Gregory 1979). Such multidisciplinary fields also include biogeomorphology and, although there is no precise definition of it, hydrological geomorphology is an area of interaction which continues to offer promising research and applied opportunities.

References

Baker, V.R. (2003) Palaeofloods and extended discharge records, in G. Benito and K.J. Gregory (eds) *Palaeohydrology. Understanding Global Change*, Chichester: Wiley.

Benito, G. and Gregory, K.J. (2003) *Palaeohydrology. Understanding Global Change*, Chichester: Wiley.

Beven, K. and Kirkby, M.J. (eds) (1993) *Channel Network Hydrology*, Chichester: Wiley.

Brookes, A. and Shields, F.D. (eds) (1996) *River Channel Restoration. Guiding Principles for Sustainable Projects*, Chichester: Wiley.

Brown, A.G. and Bradley, C. (1995) Geomorphology and groundwater: convergence and diversification, in A.G. Brown (ed.) *Geomorphology and Groundwater*, 1–20, Chichester: Wiley.

Downs, P.W., Gregory, K.J. and Brookes, A. (1991) How integrated is river basin management? *Environmental Management* 15, 299–309.

Gregory, K.J. (1979) Hydrogeomorphology: how applied should we become? *Progress in Physical Geography* 3, 84–101.

——(1996) Introduction, in J. Branson, A.G. Brown and K.J. Gregory (eds) *Global Continental Changes: The Context of Palaeohydrology*, 1–8, London: Geological Society.

——(2000) *The Changing Nature of Physical Geography*, London: Arnold.

Gregory, K.J. and Walling, D.E. (1973) *Drainage Basin Form and Process*, London: Arnold.

NRC Committee on Watershed Management (National Research Council) (1999) *New Strategies for America's Watersheds*, Washington, DC: National Academy Press.

Scheidegger, A.E. (1973) Hydrogeomorphology, *Journal of Hydrology* 20, 193–215.

Schumm, S.A. (1965) Quaternary palaeohydrology, in H.E. Wright and D.G. Frey (eds) *The Quaternary of the United States*, 783–794, Princeton: Princeton University Press.

Ward, R.C. (1966) *Principles of Hydrology*, London: McGraw-Hill.

KENNETH J.GREGORY

HYDROLYSIS

Hydrolysis is a chemical reaction of a compound with water. As opposed to HYDRATION where the water is absorbed into the compound, in hydrolysis (or 'splitting by water') both the water and the compound split up and recombine. The water is thus a reactant and not merely a solvent. For example, the reaction between a potassium-containing feldspar and water:

$$\underset{\text{feldspar}}{KAlSi_3O_8} + \underset{\text{water}}{H_2O} \rightarrow \underset{\text{silicic acid}}{HAlSi_3O_8} + \underset{\text{potassium}}{K^+} + \underset{\text{hydroxyl}}{OH^-} \quad (1)$$

The mineral releases potassium and the water splits into OH^- and H^+, the H^+ combining with the aluminosilicate from the mineral. The production of hydroxyl (OH^-) ions in solution means that the pH of the water rises. This is illustrated by grinding a mineral to powder in water and measuring the pH or 'abrasion pH'. For the more reactive minerals, this is between 8 and 11, with 8 for calcite and feldspars 8–10.

This reaction can take place in pure water at neutral pH (7). However, if the water is acidified by additional H^+ ions so the pH is below 7, the weathering reaction is accelerated. The prevailing form of acidification is by carbon dioxide:

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^- \quad (2)$$

with, for calcite:

$$CaCO_3 \rightarrow Ca^{2+} + CO_3^{2-} \rightarrow Ca^{2+} + 2(HCO_3^-) \quad (3)$$

The mineral has combined with the constituents of water, giving a free mineral ion in the water (Ca^{2+}) with one source of HCO_3^- from OH^- in the water and CO_2 and the other source of HCO_3^- from the H^+ from the water in Equation 2 combined with the CO_3^{2-} from the calcite.

Hydrolysis is thus a fundamental weathering process and it can be readily appreciated that as there are many organic sources of CO_2 through respiration and decomposition, then many such reactions are biologically originated.

STEVE TRUDGILL

HYDROPHOBIC SOIL (WATER REPELLENCY)

A soil that resists wetting by water for periods ranging from a few seconds up to days or even weeks. The reduced affinity for water is caused by a coating of long-chained organic molecules on soil particles and/or by the presence of hydrophobic (water repellent) interstitial matter. Such matter is released from a wide variety of plants through mechanical wear from leaf surfaces, decomposition of litter, release via roots and vaporization followed by condensation onto soil particles during burning, or from soil fungi and micro-organisms. The effect is temporally variable and is usually most pronounced after prolonged dry spells. Although mostly associated with semi-arid and areas with Mediterranean-type climates, it is now known to occur in a wide range of climates, including temperate and arctic-alpine environments. The potential geomorphological impacts include the restriction of soil water movement to preferential pathways; increased OVERLAND FLOW; enhanced streamflow responses to rainstorms; enhanced total streamflow; enhanced splash detachment by raindrop impact (see RAINDROP IMPACT, SPLASH AND WASH); increased SOIL EROSION by both wind and water; and increased erosion by dry creep (movement by loose, dry surface material on steep slopes). In contrast, water-repellent organic material in well-developed soil aggregates can help to stabilize them, thereby reducing soil ERODIBILITY. These impacts, however, have been largely inferred rather than demonstrated under field conditions.

Further reading

Dekker, L.W. and Ritsema, C.J. (1994) How water moves in a water repellent sandy soil. 1. Potential and actual water repellency, *Water Resources Research* 30, 2,507–2,517.

Doerr, S.H., Shakesby, R.A. and Walsh, R.P.D. (2000) Soil hydrophobicity: its causes, characteristics and hydro-geomorphological significance, *Earth-Science Reviews* 51, 33–65.

Shakesby, R.A., Doerr, S.H. and Walsh, R.P.D. (2000) The erosional impact of soil hydrophobicity: current problems and future research directions, *Journal of Hydrology* 231–232, 178–191.

SEE ALSO: fire

RICHARD A. SHAKESBY

HYPERCONCENTRATED FLOW

A flowing mixture of water and sediment transitional between a debris flow and muddy streamflow. The terms hyperconcentrated flow, hyperconcentrated flood flow, and hyperconcentrated streamflow are synonymous. The term was originally used for streamflow with sediment concentrations between 40–80 per cent by weight or 20–60 per cent sediment by volume. Rheologically the fluid appears to be slightly plastic but flows like water (Pierson and Costa 1987). Such flows are gravitationally driven, non-uniform mixtures of debris and water. They possess fluvial characteristics yet are capable of carrying very high sediment loads. Hyperconcentrated flows show clast support from grain-dispersive forces, dampened turbulence and buoyancy (implying yield strength). Sediment deposition appears to be by rapid grain-by-grain settling at the base and margins of the flow. Resultant deposits are usually either massive or display weak, near-horizontal stratification. Hyperconcentrated flows are common in volcanic environments where eruptions release large volumes of water from crater lakes or from melting of ice and snow, and when debris flows evolve downstream into hyperconcentrated flows.

Reference

Pierson, T.C. and Costa, J.E. (1987) A rheologic classification of subaerial sediment-water flows, in J.E. Costa and G.P. Wieczorek (eds) *Debris Flows/Avalanches: Process, Recognition and Mitigation, Geological Society of America Reviews in Engineering Geology* 7, 1–12.

VINCENT E. NEALL

HYPSOMETRIC ANALYSIS

Hypsometric analysis is the study of the distribution of topographic surface area with respect to altitude. The area-altitude relationship is described by the hypsometric curve (hypsographic curve) that is expressed by the function $y = f(x)$. In its absolute formulation, this curve is obtained by plotting on ordinate elevations and depths, from the top of the highest mountain to the maximum depth of abyssal trenches, and on abscissa the values of topographic surface areas. It is a cumulative curve: the abscissa of any point on it expresses the total area lying above the elevation of the corresponding ordinate.

The *absolute hypsometric curve* can be constructed for any area of land, from a small portion to the entire planet. Its use, however, is unsatisfactory when it is necessary to compare areas of different sizes and relief. To overcome this difficulty, percentage hypsometric analysis can be used, as it affords a method for expressing the area-altitude relationships in a dimensionless form (Langbein 1947).

The *percentage hypsometric curve* was used by Strahler (1952) to analyse erosional topography of drainage basins that are the basic geomorphic unit. This curve is represented by the function $y = f(x)$, but x and y are dimensionless parameters: x is the ratio of the area a above a given contour line and the whole basin area (A) and y is the ratio of the height h between the basin mouth and the contour that defines the lower limit of the

area a and the total height range in the basin (H). Obviously, x and y vary between 0 and 1. These curves can be compared irrespective of true scale as they express merely the way the landmass is distributed from base to top.

Integrating the function between the limits of $x = 0$ and $x = 1$ (or simply measuring on the graph the area under the curve) the *hypsometric integral* is obtained. It is expressed in percentage units and represents the ratio of the landmass volume of a given drainage basin to the volume of the reference solid with base equal to the basin area and height equal to the total height range in the basin. In other words the hypsometric integral measures the percentage volume of earth material remaining after the erosion of an original landmass having volume equal to the reference solid.

In their classical interpretation, Strahler's hypsometric curves and integrals identify quantitatively the stages of the Davisian geomorphic cycle. Convex curves, with hypsometric integrals higher than 0.60, indicate the inequilibrium stage of 'youth'. Smoothly S-shaped curves that cross approximately the centre of the diagram and have integrals ranging from 0.60 to 0.40 express the equilibrium stage of 'maturity' or the 'old stage'. Strongly concave curves with very low integrals result only where monadnock masses are present.

Further studies delineated another interpretation of the area-altitude analysis: the hypsometric curves express not only the stage of the 'geomorphic cycle' but also the complexity of denudational processes and the rate of the geomorphological changes in drainage basins. Such changes take place through subsequent stages of dynamic equilibrium between tectonic uplift and DENUDATION (Ciccacci *et al.* 1992); therefore each basin is marked by a hypsometric curve which is mostly a function of the denudational process type. Convex curves with a high integral refer to basins in which stream erosion is the most vigorous denudational process. Concave curves with a low integral mark basins mainly affected by intensive slope processes. Finally, hypsometric curves with an integral close to 0.5 are characteristic of basins where stream erosion balances the effectiveness of slope processes.

Actually, the classic interpretation of the hypsometric curves matches the morphodynamic characters of drainage basins in tectonically stable regions; the same interpretation is rather unsuited to explain the plano-altimetric configurations of regions affected by recent or active tectonics (Ohmori 1993; D'Alessandro *et al.* 1999).

References

Ciccacci, S., D'Alessandro, L., Fredi, P. and Lupia-Palmieri, E. (1992) Relations between morphometric characteristics and denudational processes in some drainage basins of Italy, *Zeitschrift für Geomorphology* N.F. 36, 53–67.

D'Alessandro, L., Del Monte, M., Fredi, P., Lupia-Palmieri, E. and Peppoloni, S. (1999) Hypsometric analysis in the study of Italian drainage basin morphoevolution, *Transactions Japanese Geomorphological Union* 20–23, 187–201.

Langbein, W.B. (1947) Topographic characteristics of drainage basins, *US Geological Survey, Water Supply Paper* 968-C, 125–157.

Ohmori, H. (1993) Changes in the hypsometric curve through mountain building resulting from concurrent tectonics and denudation, *Geomorphology* 8, 263–277.

Strahler, A.N. (1952) Hypsometric (area-altitude) analysis of erosional topography, *Geological Society of America Bulletin* 63, 1,117–1,142.

ELVIDIO LUPIA-PALMIERI