flotation/sinking of lighter/denser coarse clasts, respectively, under buoyancy forces.

The fifth spectrum is in the lithologies of the deposits. Pyroclastic flows are efficient conservators of heat, and so many deposits are emplaced at temperatures above those at which the juvenile material can flow plastically (e.g. >550–600 °C for rhyolitic pumice). The combination of retained heat and load stresses imposed by overlying deposits causes the juvenile fragments to adhere and flatten (weld) to form a coherent rock. At its most extreme, welding can eliminate all initial pore space and the rock may be so hot as to continue to flow plastically as a kind of lava flow. Welding can only occur as long as the juvenile phase is glassy, but in most welded deposits the glass has subsequently devitrified. In addition, gases released from the juvenile material can cause further crystallization and vapour-phase alteration of the deposit, either along discrete pathways ('fossil fumaroles') or pervasively through the porous rock mass. Non-welded deposits show little or no JOINTING, but welding (and any other causes of induration) is generally accompanied by formation of jointing in the rock mass. The orientation and spacing of the joints can vary, but columnar joints, spaced at decimetres to metres apart, are characteristic of the interior of thick ignimbrites. Closer to the base, top or sides of the deposits, or in places where local fluxes of hot gases have occurred, the jointing can be more closely spaced and fan-like in disposition.

The morphologies of freshly emplaced pyroclastic flow deposits (Figure 129) are generally very rapidly modified by erosion, as loose pyroclastic-flow material is readily eroded, generating syn- and post-eruptive debris flows, lahars and HYPERCONCENTRATED FLOWS. Incision by streams often occurs so rapidly that interaction may occur between water and the still-hot interior of the deposits, leading to 'rootless' phreatic explosions. In non-welded deposits, incision rates of metres to tens of metres per rain event are known. Incision tends to recur along the lines of the pre-eruption valleys; the greatest thicknesses of deposits (and hence the greatest compaction) occur there and so the pre-eruptive topography is mirrored in subdued fashion on the surface of the deposits, controlling the paths of re-established streams. Erosion slows considerably when hard (welded) material is reached, or the non-welded deposits are stabilized by regrowth of vegetation.

Landscape morphologies seen in areas covered by pyroclastic flow deposits reflect a complex interplay between the initial depositional morphology, the presence or absence of welding or induration to create hard rock, and the local climate. A characteristic feature in dissected large ignimbrites is a concordance of ridge or summit heights, defining a surface parallel to the original deposit surface. Slopes in non-welded deposits are typically at or close to the angle of rest, except along streams or river where undercutting leads to vertical cliffs. Slopes in welded deposits are often cliffed, as the removal of material is controlled by vertical jointing that allows toppling of columnar masses as they are undermined by erosion.

Although pyroclastic flow deposits are volumetrically important in many volcanic terrains, the enormous variety of characteristics these deposits can display, and the hazards associated with flow emplacement, mean that there is still much to be discovered about the processes and products of pyroclastic flows.

Further reading

Cas, R.A.F. and Wright, J.V. (1987) *Volcanic Successions Modern and Ancient*, London: Allen and Unwin.

Druitt, T.H. (1998) Pyroclastic density currents, in J.S. Gilbert and R.S.J. Sparks (eds) *The Physics of Explosive Eruptions*, 145–182, London: Geological Society Special Publication 145.

Fisher, R.V. and Schmincke, H.-U. (1984) *Pyroclastic Rocks*, Berlin: Springer.

Freundt, A., Wilson, C.J.N. and Carey, S.N. (2000) Block-and-ash flows and ignimbrites, in H. Sigurdsson (ed.) *Encyclopedia of Volcanoes*, 581–599, San Diego: Academic.

Ross, C.S. and Smith, R.L. (1960) *Ash-flow Tuffs: Their Origin, Geologic Relations, and Identification*, US Geological Survey Professional Paper 366.

Walker, G.P.L. (1983) Ignimbrite types and ignimbrite problems, *Journal of Volcanology and Geothermal Research* 17, 65–88.

COLIN J.N. WILSON

Q

QUICK FLOW

Hydrologists generally separate streamflow into two operationally defined components: event flow, considered to be the direct response to a given water-input event (also called direct runoff, storm runoff or stormflow), and base flow, which is water that enters from persistent, slowly varying sources and maintains streamflow between water-input events (derived largely from groundwater circulation). Quick flow is simply another term for event flow. The mechanisms involved may be one, or a combination, of Hortonian overland flow, saturation overland flow, and near-stream subsurface storm flow via groundwater mounding. In the latter case, at least some of the water identified as quick flow is 'old water' that entered the basin in a previous event. Quick flow can also be 'delayed', which involves storm runoff from distal sources via predominantly subsurface routes.

SEE ALSO: runoff generation

MICHAEL SLATTERY

QUICKCLAY

The quickclays (quick clays, quick-clays, Swedish: *kvicklera*) are clay-sized postglacial marine sediments of very high sensitivity (see SENSITIVE CLAY). The term relates to the old Nordic *queck*, meaning living. They are found in Norway, Sweden and Canada, and to a much lesser extent in Alaska, Finland and Russia, and they have been defined as having a sensitivity of greater than 50. The original definition was: a clay whose consistency changed by remoulding from a solid to a viscous fluid. Very high sensitivity values have been found – up to 200 for the Champlain clays of east Canada. The literature is dispersed; there are reviews by Bentley and Smalley (1984), Cabrera and Smalley (1973), Maerz and Smalley (1985), McKay (1979, 1982), Brand and Brenner (1981) and Locat (1995). The high sensitivity value means that the clays lose most of their strength on remoulding, and this can lead to catastrophic landslides, which progress rapidly as flowslides. Soderblom (1974) proposed that two types of quickclays should be recognized: rapid quickclays and slow quickclays. The rapid materials lose their strength very quickly on reworking; but the slow materials require the input of a fairly large amount of energy before they convert to a liquid. The strength parameters of the remoulded clays can be difficult to measure.

The classic quickclay explanation by I.Th. Rosenqvist (1953) depended on postglacial uplift, and leaching. The clay material was deposited in shallow salty seas in immediate postglacial times. As postglacial uplift occurred these deposits became dry land and were exposed to rainfall and groundwater flow. This had the effect of leaching out the salts and changing the electrochemical environment of the soil particles. The loss of the soil cations meant that the system became more metastable and responded to stress via soil structure collapse, LIQUEFACTION and flowsliding. The Rosenqvist theory appeared to work for the rapid Scandinavian clays, but not to be so suitable for the slower Canadian clays.

As mineralogical analysis became more sophisticated it became apparent that in many quickclays the actual clay mineral content was quite low and that they were perhaps better described as very fine silts. This fitted in rather well with their observed distribution on the fringes of glaciated regions. Glacial action could provide the very fine primary mineral material required to form the quickclay deposits. In fact the geomorphological observations led to a new approach to quickclays which has become known as the inactive-particle, short-range bond theory. This requires that the quickclay systems be cohesive (by virtue of the small particle size) but not plastic (because of the predominance of primary mineral particles, e.g. quartz, and the shortage of clay mineral particles). The fine blade-shaped primary mineral particles sediment in the shallow sea as Rosenqvist required, and form an open rigid structure; but the interparticle bonding is not the long-range clay mineral-type bonding but rather a short-range contact bond, enhanced by cementation.

References

Bentley, S.P. and Smalley, I.J. (1984) Landslips in sensitive clays, in D. Brunsden and D. Prior (eds) *Slope Instability*, 457–490, Chichester: Wiley.

Brand, E.W. and Brenner, R.P. (eds) (1981) *Soft Clay Engineering*, Amsterdam: Elsevier.

Cabrera, J.G. and Smalley, I.J. (1973) Quickclays as products of glacial action: a new approach to their nature, geology and geotechnical properties, *Engineering Geology* 7, 115–133.

Locat, J. (1995) On the development of microstructure in collapsible soils, in E. Derbyshire, T. Dijkstra and I.J. Smalley (eds) *Genesis and Properties of Collapsible Soils*, 93–128, Dordrecht: Kluwer.

Maerz, N.H. and Smalley, I.J. (1985) The nature and properties of very sensitive clays: a descriptive bibliography, Waterloo, Ontario: University of Waterloo Press.

McKay, A.E. (1979, 1982) *Compiled bibliography of Sensitive Clays*, Ottawa, Ontario Ministry of Natural Resources.

Rosenqvist, I.Th. (1953) Considerations on the sensitivity of Norwegian quick-clays, *Geotechnique* 3, 195–200.

Soderblom, R. (1974) New lines in quick clay research, Swedish Geotechnical Institute: Reprints and Preliminary Reports 55, 1–17.

Further reading

Ter-Stepanian, G. (2000) Quick clay landslides: their enigmatic features and mechanism, *Bulletin Engineering Geology Environment* 59, 47–57.

SEE ALSO: liquefaction; sensitive clay

IAN SMALLEY

QUICKSAND

Quicksand requires a flow of water. As the water flows through sands and silts and loses pressure its energy is transferred to the particles that it is flowing past, which in turn creates a drag effect on the particles. If the drag effect is in the same direction as the force of gravity, then the effective pressure is increased and the system is stable. In fact the soil/sediment tends to become denser. Conversely, if the water flows towards the surface, then the drag effect works against gravity, and reduces the effective pressure between the particles. If the velocity of the upward flow is sufficient it can buoy up the particles so that the effective pressure is reduced to zero. This represents a critical condition where the weight of the submerged soils is balanced by the upward-acting seepage force. This critical condition sometimes occurs in sands and silts. If the upward velocity of flow increases beyond the critical hydraulic gradient a quick condition develops.

Quicksands, if subjected to deformation or disturbance, can undergo a spontaneous loss of strength, which causes them to flow like viscous liquids. Karl Terzaghi, in 1925, explained the quicksand phenomenon as follows: first, the sand or silt concerned must be saturated and loosely packed. Second, on disturbance the constituent grains become more closely packed, which leads to an increase in pore-water pressure, reducing the forces acting between the grains. This brings about a reduction in strength. If the pore water can escape very rapidly the loss in strength is momentary. The third condition is that the pore water cannot escape readily. This occurs if the sand or silt has a low permeability or the seepage path is long, or both.

Casagrande, in 1936, demonstrated that a critical packing porosity existed above which a quick condition could be developed. He proposed that many coarse-grained sands, even when loosely packed, have porosities just about equal to the critical condition, while medium- and fine-grained sands, especially if uniformly graded (a narrow range of particle size), exist well above the critical porosity when loosely packed. Thus fine sands (say 60–150 μm) tend to be potentially more unstable than coarse-grained sands. The finer sands tend to have lower permeabilities.

Further reading

Bell, F.G. (1999) *Geological Hazards: Their Assessment, Avoidance and Mitigation*, London: Spon.

SEE ALSO: liquefaction; quickclay

IAN SMALLEY

R

RAINDROP IMPACT, SPLASH AND WASH

One of the most important driving forces in soil and hillslope EROSION is the kinetic energy of raindrops striking the soil surface. Raindrop impact contributes to soil erosion directly by splashing particles downslope, by entraining particles in OVERLAND FLOW which is below the threshold conditions necessary to pick up material. It can also affect erosion indirectly by disrupting soil aggregates, increasing ERODIBILITY, and by beating the surface into an almost impermeable seal or crust (see CRUSTING OF SOIL), which reduces infiltration and increases runoff (see RUNOFF GENERATION) discharge during rainstorms.

The kinetic energy of a moving object is expressed by $0.5MV^2$ where M = mass of the object and V = velocity. In the case of raindrops, the velocity is the terminal velocity which, in still air, reaches values around $9\,m\,s^{-1}$, for drops of 5 mm diameter (Laws 1941). During rainstorms, this value can be significantly affected by near-ground turbulence and wind. Raindrop mass is an even more critical control on the kinetic energy of raindrop impact. Raindrop size varies greatly from minute droplets a few microns in diameter, to an upper limit around 6.5 mm. As raindrop mass is directly proportional to diameter, there is a huge difference in the kinetic energy expended by small and large drops as they strike the surface. Comprehensive understanding of the relationship between raindrop impact and rainstorm characteristics is limited by the scarcity of accurate drop size measurements, particularly during rainstorms of very high intensity. However, Hudson (1981), amongst many others, has shown that rainstorms typically have a normally distributed spectrum of drop sizes, which can be expressed by the median drop diameter. This ranges from around 1.8 mm for a rainstorm of $12.7\,mm\,h^{-1}$ intensity to about 2.3 mm for a rainstorm of $65–115\,mm\,h^{-1}$ intensity. Information about characteristics of very high intensity rainfall is limited, because the most intense storms are usually of very limited duration and extent. It was thought that intensities above $150\,mm\,h^{-1}$ are very rare and largely limited to the tropics, but recent observations suggest that intensities as high as $400\,mm\,h^{-1}$ are by no means uncommon, particularly for very short periods, particularly at the beginning of thunderstorms.

Although information about raindrop size and rainfall intensities is still deficient, it is clear that there are major systematic differences between different types of rainfall and different parts of the world, which are reflected in the kinetic energy expended and the capacity of raindrop impact to generate erosion. The highest energy expenditure is certainly associated with the large drops and high intensities of severe thunderstorms or orographic rainfall and so the highest annual rainfall erosivities (see EROSIVITY) occur in areas like Assam or Hawaii, where such rainfall is combined with high annual totals. By comparison, the predominantly frontal rainfall of temperate areas produces very low kinetic energy, though occasional severe storms can, of course, cause much damage. The effect of raindrop impact is, however, strongly affected by vegetation. A dense vegetation cover can absorb virtually all the kinetic energy of raindrops, almost eliminating erosional hazard. However, although it takes about 30 m fall for drops to achieve full terminal velocity, they can achieve 60–70 per cent with a fall of some 3 m. Unless there is dense vegetation near or on the surface, raindrops can therefore regain much of their kinetic energy before hitting the surface. As a result, trees are not usually effective in controlling soil erosion in the absence of ground cover.

Raindrop impact affects the soil surface in several different ways. It may cause crusting by compacting the surface, increasing soil density and shear strength. It may also disrupt unstable soil aggregates, producing small fragments which can wash into pores and cracks, effectively sealing the surface. The resulting thin seal (often <1 mm in thickness) can make the soil surface almost entirely impermeable. The effectiveness of raindrop impact in causing compaction, disruption, crusting and sealing depends on rainfall characteristics, soil properties and on soil moisture content. Aggregate disruption by SLAKING is most effective on dry soils, while compaction is most effective on wet clay soils where cohesion drops close to zero. Although bursts of extremely high intensity rainfall, which cause most disruption, are usually very short-lived, they can strongly influence the subsequent effectiveness of erosional processes. This is particularly true in the case of intense summer thunderstorms where initial very high intensity rainfall often falls on a dry surface. These bursts usually last only a few minutes, but by initiating sealing, can result in almost instantaneous overland flow.

Raindrop impact may be entirely absorbed by soil and vegetation, but in intense storms there is usually sufficient energy available to generate some erosional processes as well. The exact processes depend on the balance between the amount of water (rainfall) arriving at the surface and the soil infiltration capacity. This will determine whether all the water can infiltrate or whether excess will be available to generate surface ponding and overland flow. Where no excess occurs, wash erosion processes are absent, but splash erosion can occur. On dry soils raindrop impact can produce miniature surface craters, but usually does not move soil particles. As the water content increases, however, soil strength drops rapidly and the surface can become fluidized. Raindrop impact is converted to an upward force which can entrain soil particles and transport them in a parabola away from the point of impact. The distance of movement depends on the mass of the particle, but is rarely more than 0.6 m above the surface, or more than 2 m in a horizontal direction, unless splash is carried by a strong wind. On a horizontal surface (in the absence of wind), movement is not significant, because the ultimate effect of many raindrops striking the surface is abundant movement, but no net transport in any direction. When the surface slopes, however, this changes as up to 60 per cent of entrained material is deposited downslope from the original impact point, so significant net transport can occur.

The relative vulnerability of soil particles and aggregates to entrainment by splash is an important component of soil ERODIBILITY. Poesen and Savat (1981) have shown in laboratory experiments that the relationship between particle size and the threshold impact energy necessary to cause entrainment is quite similar to the Hjulstrom Curve for flowing water. Entrainment of particles with diameters around 0.125 mm typically requires the lowest impact energy. Splash erosion on most slopes during most storms is therefore a selective process, which ultimately transforms the surface material, producing an *erosional lag deposit* which progressively protects the underlying soil from entrainment.

Pure splash erosion (in which material is both entrained and transported by splash) is comparatively rare, but De Ploey and Savat (1968), who originally identified the influence of slope gradient on the balance of upslope and downslope deposition of splashed material, also described the evolution of sandy hillslopes near Kinshasa, Congo, which is almost exclusively controlled by splash. Elsewhere the effects of splash erosion are subtle and often indistinguishable, but where parts of very erodible surfaces are protected by stones or bits of vegetation, the effect of splash is easily seen by the occurrence of miniature Earth pillars or hoodoos.

Splash erosion can occur without any surface water layer, but in the intense rainfall conditions which produce most splash, such a water layer usually forms quite swiftly. Initially this concentrates in micro-depressions, but ultimately it increases sufficiently in depth to overtop roughness elements and generate overland flow. Before reaching this point, however, it starts to influence the splash process. Initially, except on sandy soils, the water layer actually increases splash transport, up to a critical depth which, laboratory experiments suggest, ranges from about the diameter of the raindrops (Palmer 1963) to about one-fifth of that value (Torri *et al.* 1987). As drop size varies greatly in any rainstorm, the precise result is a very complex mixture of processes on

the surface. Eventually, however, the increasingly deep water layer protects parts of the surface from splash erosion. As the first areas protected are microtopographic depressions, the overall effect of continued splash erosion is diffusion of soil particles from higher points to these depressions, progressively reducing the amplitude of the microtopography. Another important effect is the increasing heterogeneity of soil infiltration characteristics, as the *structural* crusts which form on the high points typically have infiltration capacities up to six times higher than the *depositional* crusts which form in depressions (Boiffin and Monnier 1985).

The interaction of spatially varied rainfall, splash and microtopography produces complex, heterogeneous conditions on most hillslopes, particularly with regard to transition from splash-dominated areas to those dominated by overland flow and wash processes. On simple, idealized, homogeneous hillslopes, it is possible to distinguish an upper splash-dominated zone from a lower wash-dominated zone, and finally, from a zone in which concentrated RILL erosion occurs. In practice, the boundaries between these zones are highly irregular and dynamic. However, a transition does occur downslope as surface water deepens progressively, ultimately protecting the surface from raindrop impact. The first stages of overland flow are, however, typically very shallow. Conceptually, on very smooth surface there may actually be a thin, continuous sheet of water, but in practice as most surfaces are quite irregular, this is very rare. The initial flow usually consists of irregular, tortuous concentrations in depressions, which vary significantly in depth and width, and are separated by microtopographic protuberances. Numerous field and laboratory studies have shown that flows of this sort are typically laminar or transitional, with Reynolds numbers often well below 2,500, and relatively smooth, with Froude numbers well below 1. The flows, whether as a sheet or as more or less concentrated streams, are slow and pulsatory, and typically do not exert sufficient shear stress to entrain soil particles. However, as the Hjulstrom curve shows, flow velocities necessary to transport fine silts and clays are significantly lower than those required for entrainment. In these circumstances, raindrop impact and splash are still important, as they may be able to entrain material which can then be transported by flow. Such flows are usually referred to as *rain-impacted flows*, and the erosional process as *rainflow* or *rainwash* erosion (De Ploey 1971). The particle transport distance and the effectiveness of rainflow erosion are governed largely by particle density and settling velocity (Kinnell 2001). Significant transport is typically limited to shallow flows no more than about 1.5 times the average raindrop diameter (Kinnell 1991). Because surfaces are irregular, and flow often discontinuous, the transport distance is frequently very short, resulting in small patches of sediment deposition on the hillslope. Nevertheless, in many areas, rainflow is the most effective and frequent erosional process on upper slopes and interrill areas and can ultimately result in highly significant movement of soil to the base of the slope. This is particularly true where loose soil aggregates are of low density or are water-repellent. In some cases, the patches of sediment deposited on the slope by intermittent flows progressively join to form quite extensive *sedimentary* or *depositional seals*. These are usually highly impermeable, and become preferred locations for runoff generation and wash erosion during subsequent rainstorms (Bryan *et al.* 1978).

Once overland flow is sufficiently deep to protect the surface from raindrop impact, rainflow erosion gives way to *wash erosion*. Surface irregularities ensure that most hillslopes will have patches of wash erosion intermixed with splash and rainflow. Once the surface is fully protected, the only force which can cause entrainment is the bed shear stress exerted by flow. Transport will then occur only if shear stress exceeds the threshold necessary to move the most erodible particle. This critical value depends on soil properties, but Moore and Burch (1986) found that it was equivalent to a unit stream power of $0.002\,\mathrm{m\,s^{-1}}$ for many soils. Once unit stream power exceeds values of $0.01\,\mathrm{m\,s^{-1}}$, transport increases rapidly, and wash erosion tends to be replaced by concentrated rill erosion.

References

Boiffin, J. and Monnier, G. (1985) Infiltration rate as affected by soil surface crusting caused by rainfall, in F. Callebaut, D. Gabriels and M. DeBoodt (eds) *Assessment of Soil Surface Crusting and Sealing*, 210–217, Ghent: State University.

Bryan, R.B., Yair, A. and Hodges, W.K. (1978) Factors controlling the initiation of runoff and piping in Dinosaur Provincial Park Badlands, Alberta, Canada, *Zeitschrift für Geomorphologie, Supplementband* 34, 48–62.

De Ploey, J. (1971) Liquefaction and rainwash erosion, *Zeitschrift für Geomorphologie, Supplementband* 15, 491–496.
De Ploey, J. and Savat, J. (1968) Contribution a l'étude de l'érosion par le splash, *Zeitschrift für Geomorphologie* 12, 174–193.
Hudson, N.W. (1981) *Soil Conservation*, London: Batsford.
Kinnell, P.I.A. (1991) The effect of flow depth on sediment transport induced by raindrops impacting shallow flow, *Transactions of the American Society of Agricultural Engineers* 34, 161–168.
——(2001) Particle travel distances and bed and sediment compositions associated with rain-impacted flows, *Earth Surface Processes and Landforms* 26, 749–768.
Laws, J.O. (1941) Measurement of fall velocity of water-drops and raindrops, *Transactions of the American Geophysical Union* 22, 709–721.
Moore, I.D. and Burch, G.J. (1986) Sediment transport capacity of sheet and rill flow: application of unit stream power theory, *Water Resources Research* 22, 1,350–1,360.
Palmer, R.S. (1963) The influence of thin water layer on water drop impact forces, *International Association of Scientific Hydrology Publication* 68, 141–148.
Poesen, J. and Savat, J. (1981) Detachment and transportation of loose sediments by raindrop splash. Part II Detachability and transportability measurements, *Catena* 8, 19–41.
Torri, D. Sfalanga, M. and Del Sette, M. (1987) Splash detachment: runoff depth and soil cohesion, *Catena* 14, 149–155.

Further reading

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

RORKE BRYAN

RAINFALL SIMULATION

The purpose of a rainfall simulator is to deliver rainfall to the soil surface in a controlled manner with realistic simulation of rainfall intensity and drop-size distribution. Rainfall simulators have been used widely over the past few decades, both in the field and the laboratory. Various factors influence the method of rainfall generation including the purpose of the experiment, the soil surface area to be studied, the drop-size distribution of the simulated rainfall, the need to reproduce realistic terminal velocities, and the need for precise replication of rainfall characteristics between experiments.

Broadly, rainfall simulators fall into three categories (Foster *et al.* 2000): sprays, rotating sprays and drip-screens. Because they eject raindrops relatively high above the ground surface, spray systems are capable of achieving rainfall delivery at terminal velocities approaching that of natural rainfall. However, rainfall intensities can be hard to control, because of variation in pumping rates, and rainfall intensity usually decreases with distance from the rotating nozzle. To overcome this latter problem, multiple rotating nozzles are employed, with the overlap distance between the nozzles being determined by the area over which the simulation is to be performed (Foster *et al.* 2000). Drip systems, using hypodermic needles or drop formers, are usually used over small surface areas (typically $< 1\,m^2$). They are less likely to achieve realistic terminal velocities, because of the difficulty of raising the drip screen high enough, but give much better control of rainfall intensity. Intensities as low as $3\,mm\,hr^{-1}$ can be maintained, and replication between experimental runs is good (Bowyer-Bower and Burt 1989).

Despite the widespread use of rainfall simulators in geomorphological research, until recently there has been little co-ordinated effort to collate all the available information regarding the design and purpose of such simulators, or to discuss future developments relating to the use of this technique. To this end, the British Geomorphological Research Group established a Rainfall Simulation Working Group in 1995 to address these issues. The work resulted in a special issue of *Earth Surface Processes and Landforms* (Volume 25, Number 7, 2000) and creation of a website: http://www.geog.le.ac.uk/bgrg/index.html which includes a database of simulators and a lengthy reference list. Lascelles *et al.* (2000) make the point that when rainfall simulation is used for explicitly spatial studies, some prior analysis of the simulator's inherent variability is vital.

References

Bowyer-Bower, T.A.S. and Burt, T.P. (1989) Rainfall simulators for investigating soil response to rainfall, *Soil Technology* 1–16.
Foster, I.D.L., Fullen, M.A., Brandsma, R.T. and Chapman, A.S. (2000) Drip-screen rainfall simulators for hydro- and pedo-geomorphological research: the Coventry experience, *Earth Surface Processes and Landforms* 25, 691–707.
Lascelles, B., Favis-Mortlock, D.T., Parsons, A.J. and Guerra, A.J.T. (2000) Spatial and temporal variation in two rainfall simulators: implications for spatially explicit rainfall simulation experiments, *Earth Surface Processes and Landforms* 25, 709–721.

TIM BURT

RAISED BEACH

A raised beach is a relict depositional landform comprising mostly wave-transported sedimentary material and preserved above and landward of the active shoreline. First described by Jamieson (1908), raised beaches can form along marine coasts or lake shorelines and are well recognized as indicators of a fall in relative sea (or lake) level. In certain situations, multiple raised beaches may form adjacent to one another, producing a BEACH RIDGE plain, or strandplain (Otvos 2000). Raised beaches are distinguished here from raised marine terraces on the basis that the former are solely the product of physical depositional mechanisms, whereas the latter have a broader genesis that may incorporate depositional, erosional and/or biogenic (i.e. reefal) processes.

The elevated position of a raised beach relative to active shoreline processes may be the product of one or more of the following mechanisms: (1) tectonic uplift associated with plate-margin convergence (e.g. New Zealand east coast; Garrick 1979); (2) isostatic rebound related to ice-unloading of a land mass (e.g. mainland Scotland; Smith *et al.* 2000); (2) depositional regression involving delivery of sediment to a shoreline at a rate sufficient to allow formation and stranding of successive beaches (e.g. east coast of Australia; Thom 1984), and; (3) forced regression whereby eustatic sea-level fall leads to abandonment of a shoreline (e.g. southern Australia coast; Murray-Wallace and Belperio 1991). In the case of depositional and forced regression, the beach remains at its original elevation, as is the case for many shoreline deposits formed during the Last Interglacial sea level high-stand *c.*125 ka BP. Thus the word 'raised' is applied to all stranded fossil beaches regardless of whether the associated landmass has undergone uplift or remained stable.

Clear identification of a raised beach deposit requires satisfying a range of criteria related to the morphology and sedimentology of that deposit. Doing so allows separation from similar coastal depositional landforms such as cheniers (see CHENIER RIDGE) and linear dune ridges. For ice-free coasts, Tanner (1995) identifies four depositional processes that lead to beach ridge formation: wave-swash action, settling lag, storm surge and aeolian action. Along coasts that experience annual freeze-over of the sea (or lake) surface, ice-push is an additional mechanism for beach ridge formation. Each of these five physical mechanisms produces a shoreline deposit with different morphology and sedimentology, as described below.

The most common form of raised beach is produced by wave-swash processes on sandy to gravelly shores. Onshore transport and sorting of sediment across a beach face produces a berm that accretes to maximum wave run-up under spring tidal conditions. Subtle variations in berm morphology exist, ranging from a linear, convex-up ridge with low-angle cross-bedding to a gently landward-sloping uniform surface with continuous subhorizontal bedding. Given alongshore variations in wave energy, both forms may be present along different parts of a shoreline at one time. Consequently, it is possible to find equally variable morphology and internal structure within a raised beach.

Formation of a beach ridge by settling-lag processes is comparatively rare, developing under fetch-limited shallow water conditions such as a small lagoon or pond. Deposition occurs by sand settling out of the water column to produce a low subaqueous flat-topped ridge or bar with discontinuous horizontal bedding. Because wave action is minimal, sediments are not as well sorted as on a swash-formed beach ridge and cross-bedding is characteristically absent. Preservation of a settling-lag ridge as a raised beach typically requires a relatively rapid and permanent lowering of relative sea (or lake) level.

Storm surge is known to result in deposits at elevations above mean high water spring tidal level, either at the beach–dune interface or as a strandline feature on supratidal flats to landward of the fair-weather beach. Grain size is more varied than for fair-weather swash deposits, incorporating the largest materials available. Sedimentary structures reflect higher wave energy, ranging from complex trough cross-bedded sands to imbricated cobbles. The distinction is drawn here between these truly raised storm deposits and an overwash (see OVERWASHING) fan that is also a product of storm surge and typically located on the lagoon side of a low-lying coastal barrier, but is not raised above the elevation range of active sedimentary processes. The role of storms as an agent in the formation of raised beaches is debated in the literature (Tanner 1995), with some authors arguing for storms as an agent of net beach erosion rather than deposition. Documented instances of storm ridge formation (e.g. hurricane ridges, Florida; Tanner

1995), record these as ephemeral features, lasting only until the next storm. Good examples of multiple raised storm beaches exist along the Ross Sea coast of Antarctica where glacio-isostasy during the Holocene has driven coastal uplift (e.g. Hall and Denton 1999).

Aeolian action may contribute to the formation of a raised beach to the extent that wind-blown sand is placed directly on top of a swash-built or settling-lag initiated ridge. A raised beach with aeolian decoration is characterized by an irregular hummocky morphology with low to high-angle cross-bedding that is multidirectional and discontinuous. If vegetated, the internal structure may be weakly bedded to massive in the root zone. Relict dune ridges that are oriented parallel to a shoreline but are solely the product of aeolian processes are excluded from the range of raised beach forms.

Ice-push may also lead to formation of a beach ridge along shorelines that undergo annual freezing of the sea (or lake) surface. An ice-push ridge is typically a discontinuous accumulation of poorly sorted sand- to boulder-sized sediment that forms along the margins of winter sea-ice sheets. Ridge height is a function of the available sediment size, with boulder ridges attaining elevations of ~5 m. Due to the lack of grain sorting, the internal structure of ice-push ridges is characteristically massive. Summer wave-reworking may produce some subsequent sorting of sediment and generation of low-angle cross-bedding of the sand fraction. However, these features are mostly ephemeral, being reworked by the next ice-push.

A raised beach can be used as a proxy for palaeo-sea (or lake) level, providing the range of diagnostic physical sedimentary structures and texture noted above are preserved in the deposits. In particular, a distinction between wave-formed and aeolian sedimentary units is necessary. Thus, a vertical transition from subhorizontal or low-angle cross-bedding in medium to coarse-grained sand to high-angle cross bedding in fine to medium sand, or massive rooted structure would allow this distinction between beach berm and foredune to be drawn. Where multiple raised beaches are preserved on a strandplain, mapping of the beach-foredune contact along a dip-oriented profile can provide for reconstruction of sea (or lake) level change. Examples of this application of the raised beach sedimentary record range from decadal scale fluctuations in shoreline position along Lake Michigan (Thompson and Baedke 1995) to inferred sea level fall along the New Zealand north-east coast toward the close of the Last Interglacial period (Nichol 2002).

Where material suitable for reliable age-dating is incorporated into a raised beach deposit, it is possible to construct a chronology of formation. This is particularly useful for calculating rates of isostatic uplift (e.g. Smith *et al.* 2000), or for estimating rates of shoreline progradation in relation to local sea level and sediment supply (e.g. Tanner 1993). Traditionally, chronological analysis of raised beaches has applied radiocarbon dating to the remains of marine organisms such as shallow water molluscs (Taylor and Stone 1996). Difficulties arise with this method, however, if the material used for dating is not *in situ*. Most of the organic material incorporated in a raised beach is typically reworked from offshore environments and may therefore be considerably older than the enclosing beach sediment. Alternative dating techniques, such as optical dating of beach and dune sands, offer a more reliable avenue for establishing a detailed and accurate chronology of raised beaches, thereby enhancing their utility as a landform that can be used as an indicator of regional geomorphological and geological processes.

References

Garrick, R.A. (1979) Late Holocene uplift at Te Araroa, East Cape, North Island, New Zealand, *New Zealand Journal of Geology and Geophysics* 22, 131–139.

Hall, B.L. and Denton, G.H. (1999) New relative sea-level curves for the southern Scott Coast, Antarctica: evidence for Holocene deglaciation of the western Ross Sea, *Journal of Quaternary Science* 14, 641–650.

Jamieson, T.F. (1908) On changes of level and the production of raised beaches, *Geological Magazine* 5, 22–25.

Murray-Wallace, C.V. and Belperio, A.P. (1991) The Last Interglacial shoreline in Australia – a review, *Quaternary Science Reviews* 10, 441–461.

Nichol, S.L. (2002) Morphology, stratigraphy and origin of last interglacial beach ridges at Bream Bay, New Zealand, *Journal of Coastal Research* 18, 149–160.

Otvos, E.G. (2000) Beach ridges – definitions and significance, *Geomorphology* 32, 83–108.

Smith, D.E., Cullingford, R.A. and Firth, C.R. (2000) Patterns of isostatic land uplift during the Holocene: evidence from mainland Scotland, *Holocene* 10, 489–501.

Tanner, W.F. (1993) An 8000-year record of sea level change: data from beach ridges in Denmark, *Holocene* 3, 220–231.

Tanner, W.F. (1995) Origin of beach ridges and swales, *Marine Geology* 129, 149–161.

Taylor, M. and Stone, G.W. (1996) Beach-ridges: a review, *Journal of Coastal Research* 12, 612–621.

Thom, B.G. (1984) Transgressive and regressive stratigraphies of coastal sand barriers in eastern Australia, *Marine Geology* 56, 137–158.

Thompson, T.A. and Baedke, S.J. (1995) Beach-ridge development in Lake Michigan – Shoreline behaviour in response to quasi-periodic lake-level events, *Marine Geology* 129, 163–174.

SEE ALSO: beach ridge; chenier ridge; sea level; strandflat

SCOTT NICHOL

RAMP, COASTAL

The term 'ramp' has been used by some workers to refer to gently sloping SHORE PLATFORMS, particularly to those in the north Atlantic, in order to distinguish them from the subhorizontal platforms which are more common in Australasia. Generally, however, the term is either used for sections of higher gradient at the rear of gently sloping shore platforms, or for steeply sloping rock surfaces (commonly 4° to 10°) that occupy the entire intertidal zone and may extend to elevations that are well above the high tidal level. Both types of ramp have been reported most frequently from the swell wave environments of Australasia and elsewhere around the Pacific, and less frequently from the storm wave environments of the mid-latitudes of the northern hemisphere. It has been suggested that ramp occurrence and morphology are related to the strength and frequency of the swash generated by storm waves, to waves of translation that sweep across the platforms, and to the presence of abrasive material at the cliff foot. In northeastern England, ABRASION accomplishes rapid erosion, ranging up to 30 mm yr^{-1}, on the steeply sloping ramp where there is a sand and pebble beach, whereas dessication of the shale is dominant on the more gently sloping platform (Robinson 1977). In some places the occurrence of ramps appears to reflect variations in rock structure and lithology. The presence of thick shale beds and other weak material near the high tidal level seems to be particularly suitable for the development of prominent ramps in eastern Canada and in northeastern England, and this is supported by mathematical modelling, which suggests that ramps are most common where rapid erosion produces wide intertidal platforms. Where contemporary rates of erosion are low, however, as in northwestern Spain, ramps extending up to several metres above the modern high tidal level are the result of higher SEA LEVEL during the last interglacial (see ICE AGES) (Trenhaile *et al.* 1999). In southern Australia, sloping ramps, which extend up to more than 10 m above present sea level, are probably polygenic, having developed under rising and falling sea level during the Cenozoic Era (Young and Bryant 1993).

References

Robinson, L.A. (1977) Erosive processes on the shore platform of northeast Yorkshire, England, *Marine Geology* 23, 339–361.

Trenhaile, A.S., Pérez Alberti, A., Martínez Cortizas, A., Costa Casais, M. and Blanco Chao, R. (1999) Rock coast inheritance: an example from Galicia, northwestern Spain, *Earth Surface Processes and Landforms* 24, 605–621.

Young, R.W. and Bryant, E.A. (1993) Coastal rock platforms and ramps of Pleistocene and Tertiary age in southern New South Wales, Australia, *Zeitschrift für Geomorphologie* 37, 257–272.

ALAN TRENHAILE

RAPIDS

Rapids in bedrock channels are not technically defined in fluvial literature, but imply steep reaches with rough water and very variable depth between lower gradient pools (Leopold 1969). Their origin is attributed to the erratic and episodic supply of boulders into the channel, both debris flows from tributaries and rock avalanches and rock fall from the valley sides (Howard and Dolan 1981; Webb *et al.* 1984). Subsequent accelerated flow through the constriction redistributes boulders downstream and partly reshapes the channel bed into quasi-stable form of boulder-strewn bars (Graf 1979; Kieffer 1987).

References

Graf, W.L. (1979) Rapids in canyon rivers, *Journal of Geology* 87, 533–551.

Howard, A.D. and Dolan, R. (1981) Geomorphology of the Colorado River in the Grand Canyon, *Journal of Geology* 89, 269–298.

Kieffer, S.W. (1987) The rapids and waves of the Colorado River, Grand Canyon, Arizona, Report 87–096, *United States Geological Survey*.

Leopold, L.B. (1969) The rapids and the pools – Grand Canyon, *United States Geological Survey Professional Paper* 669-D, 131–145.

Webb, R.H., Pringle, P.T., Reneau, S.L. and Rink, G.R. (1984) Monument Creek debris flow, 1984: implications for formation of rapids on the Colorado

River in Grand Canyon National Park, *Geology* 16, 50–54.

KEITH J. TINKLER

RASA AND CONSTRUCTED RASA

The term rasas, of Spanish derivation, refers to old and perched littoral levelling surfaces or planation surfaces. Their width can reach several kilometres. The erosion surfaces are bordered inland by steep relief and by cliffs towards the sea. They were described for the first time by Hernandez-Pacheco (1950) on the Cantabrian Coast, northern Spain. Guilcher (1974) made a remarkable synthesis. These forms were also observed in Galicia (Nonn 1966), northern Chile (Paskoff 1970), southern Morocco, Brittany and Cornwall in England (Guilcher 1974), and Sardinia (Ozer 1986). Guilcher distinguished three types of rasas. The first one was described above, the second is more complex and is constituted by a succession of levellings arranged in stairs, and the third is when the passage towards the inland is gradual.

Plate 93 Rasa: Coast of Gallura (north Sardinia)

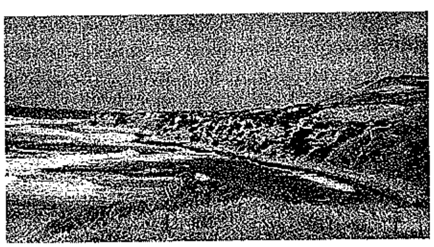


Plate 94 Constructed rasa: Coast of Anglona (north Sardinia). Accumulation of aeolianites on the terrace of the last interglacial sea level

Many of these rasas are covered by marine deposits (sand and rounded pebbles). These sediments were brought at a later date, during tertiary transgressions which only slightly retouched these levelling surfaces. Evidence of this process is found through ancient reefs in Brittany (Guilcher 1974), northern Sardinia (Ozer 1986) and south of Tangier, Morocco (Ozer, 2001 observation).

However, a convergence of shapes can exist, which is then called constructed rasas. This is a littoral aeolian accumulation, generally indurated (aeolianites), often mixed with local deposits of torrential origin. These accumulations are cut again in a shelf shape, slightly sloping towards the sea subsequent to runoff erosion.

The most spectacular constructed rasas are developed on slopes preceded by a well-developed continental shelf exposed to dominant winds. During Quaternary regressions, winds transported abandoned sands from the continental shelf until the first relief was formed by ancient cliffs which developed during the Quaternary transgressions. These deposits, essentially aeolian, became consolidated and were later shaped into cliffs by the current sea level. They are bounded inland by strong relief which is a previous Quaternary dead cliff.

References

Guilcher, A. (1974) Les 'rasas': Un problème de morphologie littorale générale, *Annales de Géographie* 455, 1–32.

Hernandez-Pacheco, E. (1950) Las rasas litorales de la costa cantabrica en su segmento asturiano, *C.R. Congrès Internationnal Géographie de Lisbonne* 2, 29–86.

Nonn, H. (1966) *Les régions côtières de la Galice (Espagne), étude morphologique*, Paris, Strasbourg: Thèsè.

Ozer, A. (1986) Les niveaux marins au Pléistocène supérieur en Méditerranée occidentale, Atti del Convegno 'Evoluzione dei litorali', ENEA, Policoro (Italia), 241–261.

Paskoff, R. (1970) *Recherches géomorphologiques dans le Chili semi-aride*, Bordeaux: Thèse.

ANDRÉ OZER

RATES OF OPERATION

Rates of operation of geomorphic processes are determined in a number of different ways depending on the time and space scales of interest, and whether one is interested in rates of operation of individual processes or in the aggregate rates resulting from all processes combined.

The current dynamic tectonic conditions need to be considered alongside of the overall denudation rates in order to place measurement programmes conducted at site or watershed scale into proper perspective (Brunsden 1990). Brunsden notes that with respect to the major geotectonic provinces the Cenozoic orogenic regions, and especially the subduction areas on plate margins, can experience greater than 20 $mm\,yr^{-1}$ of vertical movement at the same time as the overall denudation rate rarely exceeds 1 $mm\,yr^{-1}$. At the other extreme, shields, platforms, cratonic regions and intracratonic basins experience less than 1 mm/1,000 yrs of vertical movement; nevertheless, overall denudation rates scarcely exceed 1 mm/10,000 yrs. Superimposed on these orogenic and epeirogenic movements are the isostatic readjustments which occur in regions recently emerged from under thick ice sheet cover. In the cratonic regions of the Baltic Sea and Hudson Bay, rates of isostatic readjustment were as high as 1–10 m/100 yrs at the close of the Wisconsinan glaciation and remain as high as 10 $mm\,yr^{-1}$ in the Gulf of Bothnia. The implication drawn from these data corresponds closely with that of Schumm (1963) namely that 'the style and location of landform change is determined by the type, location and rate of tectonic movements and their associated stress fields over the relevant time and space framework of the landform assemblage' (Brunsden 1990: 3). There is general agreement on the order of magnitude of these rates at global to regional scale; the extent to which they are relevant to site and watershed scale is open to debate.

Average rates of operation of processes can obscure the fact that many processes are episodic and that land surfaces may evolve in a series of step jumps, with periods of relative stability followed by brief periods of rapid erosion or accelerated uplift. Variations in rates of change through time are further complicated by variations in space. Even within geotectonic provinces, spatial variations can be large.

Fundamentally, landscape stability and rates of change depend upon the ratio of resistances to change to the forces promoting change. Where these forces are in balance, little change occurs; where resistance exceeds the forces of denudation, weathering processes permit the deepening of soil profiles. This condition is called transport-limited. Where the forces of denudation are greater than the landscape resistance, erosion removes soil and weathering products as quickly as they are formed. This condition is called weathering-limited.

It is apparent that erosion rates will depend in large measure on the availability of transportable soil and sediment. As soil can only accumulate to considerable depths under stable conditions and eroded sediment can only accumulate at regional scale under conditions of continental-scale glaciation, the most extreme erosion rates occur when there is a marked change from one set of processes to another. When a threshold between one set of processes and another is crossed, extremely high rates of denudation may occur. The time period over which these accelerated rates can last is limited by the supply of readily eroded soil and sediment. Paraglacial geomorphology is one striking example of accelerated erosion and sedimentation following threshold exceedance. Landscape sensitivity to change is therefore as effective in controlling the short-term denudation rate as is the energy of the processes of erosion and transport.

In a brief historical sketch of the development of interest in rates of operation of geomorphic processes, Archibald Geikie and Charles Darwin are two of the early researchers who attempted to determine the rates of operation of individual processes. Geikie estimated the rate of rock weathering by measuring changes on dated tombstones in Edinburgh churchyards and Darwin estimated the rate of soil movement on slopes caused by worm casting. The first spatially representative estimates of the overall rate of ground loss derived from a summary of river sediment loads in the United States. Early twentieth-century estimates of the rate of cliff retreat in Germany on sandstones and in Brazil on granites under rainforest found that the rates in Brazil were an order of magnitude greater than those in Germany. Seasonal rates of movement of stones on talus in the Alps and longer term integrations of postglacial creep of till (135 m in 30,000 years) and Lester King's estimates of the rate of retreat of the Drakensberg scarp in South Africa (240 km in 150 million years) were some of the few quantitative rates of erosion estimated before the 1950s. No one seems to have correlated these data as they were simply too scattered and lacking in formal methodology. One notable exception was the US Soil Conservation data. The first systematic programme to measure soil erosion came about in the United States during the 1930s when one of the New Deal programes of President Roosevelt, intended to stem the growth of unemployment, resulted in the construction of tens of thousands of small dams by the US Soil Conservation Service. Large data sets of volumes of sediment delivered to small reservoirs thereby became available. Accelerated erosion plots usually included an adjacent control plot to demonstrate the negative effects of poor land use practices. From a strictly geomorphic perspective, the control plot data gave indications of spatial variability of surface wash rates, but integrated analyses were not published until the late 1940s. One of the important theoretical contributions from the US Soil Conservation data was the formulation of the dynamic concept of sediment sources. There was a recognition of the difference between sediment sources and sediment delivery at the outlet of each basin and the sediment delivery ratio became a useful tool to determine sediment storage. The 1950s were a decade of pioneering studies on rates of geomorphic process, all the way from sediment budgeting (Jackli 1957; Rapp 1960; Leopold *et al.* 1966) to surface wash (Schumm 1956) and a variety of creep processes (Jahn 1961).

By 1983, Saunders and Young summarized (somewhat uncritically) literally thousands of reported data on rates of process operation. Data are no longer the problem but standardized data, both in terms of methods of collection and units of measurement, remain a serious problem. With respect to endogenic processes, England and Molnar (1990) summarized the major difficulties. Many reports of surface uplift in mountain ranges are based on mistaking exhumation of rocks or uplift of rocks for surface uplift and provide no information whatsoever on the rates of surface uplift. Some observations provide reliable measures of the uplift of rocks but, because erosion rates may be high, the mean surface elevation may be decreasing while the rocks are uplifting.

Standardization of data

How does one compare (a) the linear downslope movement of the uppermost layer of the regolith with (b) the volumetric downslope movement of the whole regolith with (c) the slope retreat or ground loss perpendicular to the ground surface with (d) the mass of sediment transported past a control section with (e) the bedrock mass uplifted above the geoid surface? These are all common ways of reporting the results of contemporary process measurements. Caine (1976) stated the problem coherently. Not only is there a problem of the use of disparate units and dimensions, but there is a need to define hillslope erosion and river channel erosion in terms that are mutually compatible, and storage effects within river systems should also be accounted for. His solution is the calculation of a unit of geomorphic work which incorporates the product of the mass of sediment, the change in elevation and the gravitational acceleration. The approach is logically compelling but has not been widely adopted.

An alternative solution has been to convert all data to a linear measure of denudation distributed evenly across the basin. The Bubnoff unit (Fischer 1969), which is equivalent to 1 mm of denudation per 1,000 years, has also encountered some resistance, partly on account of the somewhat arbitrary specific gravity and packing corrections that have to be made, but also because of the impression created of even denudation across a highly spatially variable surface. It seems fair to say that the prevailing attitude is to maintain different units of measurement for slope, channel and basin data.

Equilibria between hillslope erosion and sediment yield

A number of studies have engaged the question of the quantitative balance between hillslope erosion or contemporary uplift and sediment yield. Here we consider just two examples of apparent balance between measured rates. Adams (1980) examined the Southern Alps of New Zealand and compared rates of crustal shortening, tectonic uplift, river sediment and dissolved load, and offshore deposition. In billions of $\mathrm{kg\,yr^{-1}}$, the rates were respectively of the order of 700, 600, 700 and 580. Data on crustal shortening derived from geophysical estimates of the rate of convergent plate motion across the Indian–Pacific plate boundary, amounted to about $22\,\mathrm{mm\,yr^{-1}}$. This process would lead to a build-up of crustal lithosphere. Data on tectonic uplift were calculated by converting the shortening to uplift along the Alpine Fault. Data on river loads were taken from water analyses (dissolved load), estimates from formulae and field measurement (bedload) and monitored data supplemented by runoff vs sediment concentration relations (suspended load). The average amount removed was adequate (on an annual basis) to balance the build-up effect from tectonic uplift. Finally, data on offshore deposition showed similar order of magnitude effects, thereby removing sediment to the east and west to the converging plate margins. The model described by Adams is a steady-state mountain range with rapid uplift being balanced by rapid erosion. The details are contentious, but the example is instructive in that it demonstrates the extensive data demands placed on such an interpretation. The author is fully cognizant of the errors inherent in the calculations. He confirms his findings in an interesting appeal to the shapes of New Zealand's mountains. The Southern Alps are spiky mountains (suggesting a steady-state condition) whereas immediately adjacent, in Otago, the mountains are flat-topped and are the remains of a pre-uplift surface of low relief.

Reneau and Dietrich (1991) examined a part of the southern Oregon Coast Range and compared data on bedrock exfoliation rates, thicknesses and dates of accumulations of colluvial fill in topographic hollows and the size of the contributing source area with monitored suspended and dissolved load data from the region. The novelty of this approach derives from some premises with respect to the effectiveness of topographic hollows in trapping colluvium and the ability to satisfactorily date the colluvial fill at up to five stratigraphic levels. If it be admitted that colluvial transport rates down the axis of a hollow are dependent on gradient and are constant in the part being evaluated, then net deposition is entirely due to colluvium added from the adjacent side slope. Calculations of volumetric colluvial transport rates into each hollow involved using measures of local topographic convergence, average soil density and the mass depositional rate of colluvium. Calculated average erosion rates from dated hollows were equivalent to about 70 Bubnoffs (mm/1,000 yrs); calculated exfoliation rates were equivalent to about 90 B and calculated denudation rates varied from 50–80 B. Again, the authors carefully identify error bars on their data but conclude that because hillslope and basin-wide erosion rates are so similar hillslope sediment production and stream sediment yield in the Oregon Coast Range are roughly in balance. Net changes in sediment storage downstream are necessarily also minor. Again it should be noted that the data needs are onerous and creative field measurement programmes are necessary.

By contrast with rates of geomorphic process in apparently steady-state environments, relatively few measurement programmes on rates of bedrock incision have been reported.

Whipple *et al.* (2000) took advantage of the diversion of the upper Ukak River in Alaska by an ash flow in 1912 to measure rates of incision along a newly formed bedrock channel. Although the minimum rates of incision are high (10–100 $\mathrm{mm\,yr^{-1}}$), they are within the range of previously published estimates (e.g. Stock and Montgomery 1999). In this branch of process geomorphology there are substantially more modelled rates of operation of process than confirmed field data.

There remains considerable ambiguity over the significance of measured rates of erosion at site scale and over short periods of time vis-à-vis the evolving shape of the landscape. During the 1960s there was optimism that measured rates might be extrapolated from site scale and from short-term measurements to larger landscapes and longer term rates. Such expectations have been shown to be naive and derived in part from assumptions about equilibrium, a balanced condition and the ignoring of contingent environmental constraints. Perhaps the central question now being engaged is that of how to link measurements of rates of operation of geomorphic process at one scale (whether temporal or spatial) to another scale. The information is urgently required in the context of concerns about global environmental change (at what scales are the effects of human activity clearly differentiable from the effects of climate change?) and also in the context of a better understanding of Earth history.

References

Adams, J. (1980) Contemporary uplift and erosion of the Southern Alps, New Zealand, *Geological Society of America Bulletin* 91, 1–114.

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

Caine, N. (1976) A uniform measure of subaerial erosion, *Geological Society of America Bulletin* 87, 137–140.

England, P. and Molnar, P. (1990) Surface uplift, uplift of rocks and exhumation of rocks, *Geology* 18, 1,173–1,177.

Fischer, A.G. (1969) Geological time-distance rates: the Bubnoff unit, *Geological Society of America Bulletin* 80, 549–552.

Jackli, H. (1957) Gegenwartsgeologie des bundnerischen Rheingebietes: ein beitrag zur exogenen Dynamik Alpiner Gebirgslandschaften, *Beiträge Geologie Schweiz Geotechnische Serie*, No. 36.

Jahn, A. (1961) Quantitative analysis of some periglacial processes in Spitzbergen, *Panstwowe Wydawnictno Naukowe*, Warsaw, Geophysics, Geography and Geology, IIB.
Leopold, L.B., Emmett, W.W. and Myrick, R.M. (1966) Channel and hill slope processes in a semi-arid area, *US Geological Survey Professional Paper* 352-G. Washington, DC: US Geological Survey.
Rapp, A. (1960) Recent development of mountain slopes in Karkevagge and surroundings, northern Scandinavia, *Geografiska Annaler* 42A, 65–200.
Reneau, S.L. and Dietrich, W.E. (1991) Erosion rates in the southern Oregon Coast Range: evidence for an equilibrium between hill slope erosion and sediment yield, *Earth Surface Processes and Landforms* 16, 307–322.
Saunders, I. and Young, A. (1983) Rates of surface processes on slopes, slope retreat and denudation, *Earth Surface Processes and Landforms* 8, 473–501.
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.
——(1963) The disparity between present rates of erosion and orogeny', *US Geological Survey Professional Paper* 454-H, Washington, DC: US Geological Survey.
Stock, J.D. and Montgomery, D.R. (1999) Geologic constraints on bedrock river incision using the stream power law, *Journal of Geophysical Research* 104, 4,983–4,993.
Whipple, K.X., Snyder, N.P. and Dollenmeyer, K. (2000) Rates and processes of bedrock incision by the Upper Ukak River since the 1912 Novarupta ash flow in the Valley of Ten Thousand Smokes, Alaska, *Geology* 28, 835–838.

Further reading

Burbank, D.W. and Beck, R.A. (1991) Rapid, long-term rates of denudation, *Geology* 19, 1,169–1,172.
Gage, M. (1970) The tempo of geomorphic change, *Journal of Geology* 78, 619–625.

SEE ALSO: Bubnoff unit; chemical denudation; denudation

OLAV SLAYMAKER

REDUCTION

Reduction is the gain of a negative electron so an element becomes less positively charged, for example ferric iron, Fe^{3+} (or Iron III) becomes reduced to ferrous iron Fe^{2+} or Iron II. This process commonly occurs in the absence of oxygen but can equally occur when iron is in an acid solution. The latter process accounts for the solubilization and loss of iron oxides in the upper parts of soil profiles where there is, in fact, oxygen available but where organic acids derived from the decomposition of plant material acidifies the soil.

In geomorphology, the focus is on the mobilization of iron under reducing conditions and its transport in anoxic/acid waters, often in deep ground water, and the redeposition of iron oxides in oxic conditions. Retallack (1992) proposes that a study of fossil soils shows how the Earth's atmosphere evolved with the gradual increase in oxygen due to the rise of the plants. Around 1,000 million years BP virtually all palaeosols contained oxidized iron, but palaeosols with reduced iron present occur before that date and 3,000 million years ago there were very few paleosols with oxidized iron present.

Reference

Retallack, G.J. (1992) *Soils of the Past*, London: Unwin Hyman.

STEVE TRUDGILL

REEF

Reefs can broadly be classified as spatially heterogeneous, three-dimensional structures which have morphological form that is different from that of the underlying substrata. Historically and currently the term reef has been used to classify a whole host of organic and inorganic structures including stone reefs, OYSTER REEFS, CORAL REEFS, ATOLLS, SERPULID REEFS, algal reefs and artificial reefs. Due to the range of disparate features being classified as reefs, there has been much debate in the literature over what does and does not constitute a reef.

Although many reef specialists, both biologists and geologists, have argued that reefs must be of biogenic origin to be classified as reefs, numerous applications of the term reef have been applied to inorganic structures. In the late nineteenth and early twentieth centuries, emminent natural scientists and geologists referred to inorganic structures, such as beachrock or the curious bar at Pernambuco, Brazil, as stone reefs (e.g. Branner 1905). More recently, there has been a resurgence of the use of the term reef for inorganic structures, as artificial reefs. In many coastal environments, artificial reefs are being built for a range of purposes including as offshore coastal defences, such as those at Sea Palling, Norfolk, England; or as subtidal structures designed to enhance biodiversity in inshore waters. From a geomorphological perspective, both organic and inorganic reef structures influence geomorphological processes and the morphology of coastal environments. As such, both organic and inorganic reef structures are classified as reefs for geomorphological research purposes.

Smaller reefs have often been termed bioherms and biostromes: bioherms are reef-like, mound-like or lens-like features of purely organic origin which are found embedded in rocks of different lithologies, while biostromes are organic layers, which are thinner and less developed structures than bioherms, such as oyster reefs (Cummings 1932).

Importantly, Cummings was one of the first authors to stipulate that reefs are organic forms which can be produced by several different species and they exhibit a variety of forms, ranging from reefs to bioherms and biostromes, where corals are only one type of reef form. Although this subdivision of reefs into more specialized categories had many merits, modern authors still preferentially use the term reef.

Reefs are found in temperate to tropical marine ecosystems, with the most prominent reef types, corals and atolls, being found in tropical and subtropical zones. Algal reefs and bioherms are commonly found in more moderate climatic zones, such as the Mediterranean, and include corniches, trottoirs and mini-atolls built primarily by calcarous algae, vermetids and serpulids. In temperate regions, reefs are often more like bioherms or biostromes in structure and include reef communities such as *Sabellaria*, oyster or skeletal carbonate reefs. Temperate reefs are found in the eulittoral to pelgaic zones and typically develop on a firm substrata. Reefs can enhance the growth and persistence of other species, by providing sheltered habitat or by providing a fixed substrata upon which cryptic communities can colonize and they can also influence sediment dynamics by trapping and storing sediment.

References

Branner, J.C. (1905) Stone reefs on the northeast coast of Brazil, *Geological Society of America Bulletin* 16, 1–12.
Cummings, E.R. (1932) Reefs or bioherms? *Geological Society of America Bulletin* 43, 331–352.

Further reading

Fagerstrom, J.A. (1987) *The Evolution of Reef Communities*, New York: Wiley.
Riding, R. (2002) Structure and composition of organic reefs and carbonate mud mounds: concepts and categories, *Earth-Science Reviews* 58 (1–2), 163–231.

Wood, R. (1993) Nutrients, predation and the history of reef-building, Palaios 8, 526–543.

LARISSA NAYLOR

REGELATION

Means to 'freeze again'. In the glacial context it refers to those processes which permit a glacier to slide over a rough bed by means of melting on the upglacier side of an obstacle and to refreeze on the downglacier side. Regelation occurs because the greatest resistance to glacier movement is on the upstream side of an obstacle. This results in locally high pressures and a consequent lowering of the pressure melting point. Thus melting of ice occurs immediately upglacier of the obstacle, and the resulting meltwater migrates to the lower pressure zone on the downglacier side of the obstacle. There it refreezes because the pressure melting point is higher. It is through this mechanism that the ice in effect overcomes the obstacle by temporarily turning to water and back again. It is, therefore, an important process in glacier sliding and has been confirmed by direct observation in subglacial cavities.

Further reading

Weertman, J. (1957) On the sliding of glaciers, *Journal of Glaciology* 3, 33–38.

A.S. GOUDIE

REGOLITH

The term was coined by Merrill (1897) to describe an 'incoherent mass of varying thickness composed of materials essentially the same as make up the rocks themselves, but in greatly varying conditions of mechanical aggradation and chemical combination'. He went on to point out regolith may be formed *in situ* or from sediments transported from another source.

Merrill derived the word from the Greek regos ($\rho\eta\gamma o\sigma$) meaning blanket or cover and lithos ($\lambda\iota\theta o\sigma$) meaning rock or stone. Jackson (1997) defines regolith as a term for 'the layer or mantle of fragmental and unconsolidated material, whether residual or transported and highly varied in character, that nearly everywhere forms the surface of the land and overlies the bedrock'. Another more simple definition is everything that lies between fresh rock and fresh air.

Regolith is restricted to terrestrial environments and is generally considered to comprise mechanically and chemically weathered rock debris whether *in situ* or transported. It includes rock weathered to varying degrees, sediments of colluvial, alluvial, aeolian, marginal marine and glacial origin as well as volcanic ash and lag gravels, pisolites and sand. It ranges from soft and loose to consolidated and/or cemented and very hard.

Regolith is the earth material usually called 'soil' by many scientists and engineers. Engineers tend to call any earth materials that can be moved with a bulldozer or mechanical digger soil. Forensic geologists call their sampling media soil. Agricultural scientists on the other hand think of soil as a growing medium from crops and pasture. To a regolith scientist soil is a part of the regolith at the uppermost part of the whole body of unconsolidated material they call regolith. Regolith also may contain buried soils that formed during periods when little accretion occurred in an accretionary (sedimentary or volcanic) landscape.

Both Merrill and Jackson consider regolith to be unconsolidated, but when duricrusts are considered this concept falls down. A silcrete for example is as tough a rock as one can find, but it is considered by most to be part of the regolith. Equally in many parts of the world lava flows are encapsulated by regolith (Figure 130). Does this mean that the lavas are part of the regolith or are there two different regolith units above and below the lava? In Figure 130 it is clear that at section A this is the case, but laterally in the section there is only one regolith, part a lateral equivalent of the one below the lava and one laterally equivalent to that above.

This dilemma raises the issue of regolith stratigraphy and dating. Within transported parts of regolith it is possible to apply the principles of lithostratigraphy remembering that this provides little in the way of chronological control on regolith materials. The lithostratigraphy in section A (Figure 130) is very different from that in section B. The age of weathering in the regolith unit below the lava may very well be very different from that above it. The age of weathering in section B will be complex because this section has been exposed to weathering for a longer time than the upper regolith unit in section A and probably has a complex weathering profile carrying components of pre- and post-lava weathering. Moreover because weathering occurs continuously, albeit at different rates, weathering overprints on regolith materials cannot be used for correlation unless dating of weathering demonstrates equivalence. Without dates on regolith materials this dilemma cannot be resolved except in a relative sense, and even then with some difficulty.

The age of regolith does not form part of its definition, but many would consider that Palaeozoic (or even older) materials now at the surface are not regolith but exhumed surfaces on which some regolith is preserved. In many parts of the world ancient regolith exists at or near the modern surface. In some cases it is unlikely that these surfaces and materials have ever been buried (Craig and Brown 1984). In other cases they were buried and have since been exhumed (Lidmar-Bergström 1995). Carboniferous weathering profiles have been dated by palaeomagnetic methods within 1–2 m of the surface in central New South Wales, Australia (Pillans *et al.* 1999), but it has been suggested this profile is exhumed several times, 3.5 km during the Permo-Carboniferous and another 2.5 km during the Triassic to early (O'Sullivan *et al.* 2000). Most regolith however is very much younger than the Palaeozoic and it is still forming across the Earth's surface.

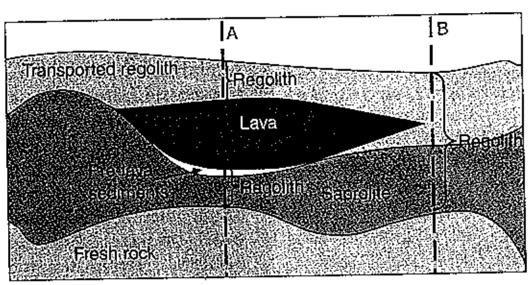


Figure 130 Is it logical to include the lava in with regolith or define it as detached and possibly part of the fresh rock even though it may be a different age and composition?

In situ regolith generally forms a weathering profile and these profiles often have a characteristic sequence of materials developed in them. Taylor and Eggleton (2001) provide detailed descriptions and interpretations of weathering profiles. Essentially the sequence is:

- soil
- ferruginous and/or aluminous lag
- collapsed saprolite (may be mottled by ferric oxihydroxides)
- saprolite mottled by ferric oxihydroxides
- bleached saprolite (composed of kaolinite and/or quartz grading downward into more complex clay minerals and quartz ± other primary minerals)
- saprock
- weathering front
- fresh rock.

Weathering profiles of this type are often considered to be the norm and if the upper parts of the profile (e.g. the lag and/or collapsed saprolite) are not present it is often inferred that there has been erosion. This is a misguided inference as there is often no evidence to suggest that the profile was completely developed or that it ever had all those components. Such inferences can lead to erroneous conclusions regarding landscape evolution and the formation of various regolith materials.

References

Craig, M.A. and Brown, M.C. (1984) Permian glacial pavements and ice movement near Moyhu, north-east Victoria, *Australian Journal of Earth Sciences* 31, 439–444.

Jackson, J.A. (1997) *Glossary of Geology*, 4th edition, Alexandria, VA: American Geological Institute.

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

Merrill, G.P. (1897) *A Treatise on Rocks, Rock Weathering and Soils*, New York: Macmillan.

O'Sullivan, P.B., Pain, C.F., Gibson, D.L. *et al.* (2000) Long-term landscape evolution of the Northparkes region of the Lachlan Fold Belt, Australia: constraints from fission track and paleomagnetic data, *Journal of Geology* 108, 1–16.

Pillans, B., Tonui, E. and Idnurm, M. (1999) Palaeomagnetic dating of weathered regolith at Northparkes Mine, N.S.W., in G. Taylor and C.F Pain (eds) *Regolith '98: New Approaches to an Old Continent*, 237–242, Perth: CRC LEME.

Taylor, G. and Eggleton, R.A. (2001) *Regolith Geology and Geomorphology*, Chichester: Wiley.

GRAHAM TAYLOR

REJUVENATION

Rejuvenation stems from *juvenis*, Latin for young. Thus rejuvenation is to make young again. The term has been applied to individual landforms such as a hillslope or a river channel, but it is most commonly and more appropriately applied in the context of the entire landscape. The term enjoys wide usage among physical geographers and historically based geomorphologists. Its origin and usage in geomorphology can be traced to the interpretation of several lengthy philosophical discourses in the late nineteenth century when some of the major paradigms of long-term landscape evolution were first established (Davis 1889, 1899).

The geographic cycle of Davis (1899) (see CYCLE OF EROSION) continues to influence modern thoughts on long-term landscape evolution. Davisian theory explains landscapes and their constituent landforms primarily in the context of the amount of time that they have been subjected to the forces of erosion. Landscapes are viewed as being born from impulsive rock uplift above sea level. This uplift is followed by a protracted period of erosion that lowers the mean elevation of the landscape by first incising deep, narrow valleys, then widening the valley bottoms and rounding the hillslopes, finally leading to the decline of interfluves to the point that the entire landscape has been reduced to a flat plain or PENEPLAIN. During the valley incision stage, the landscape is traditionally described as youthful, in the valley widening and hillslope rounding phase, the landscape is thought of as mature, and as a peneplain, the landscape is thought of as old. Davis (1899), as well as the subsequent generation of geomorphic thought, recognized that in reality, the geographic cycle almost never proceeded to completion creating a widespread peneplain. Rather, tectonism was understood to be frequent enough such that landscapes in various stages of maturity or old age were uplifted, increasing mean elevation, causing renewed stream incision, and effectively making the landscape appear young again. Such active tectonics has the effect of rejuvenating the landscape.

Rejuvenation is a useful concept when viewing landscape evolution over long (10^6–10^7 yrs) timescales, especially when the flux of sediment that is eroded from those landscapes is considered (Schumm and Rea 1995). Long-term sediment yield from landscape erosion tends to follow a decaying exponential relationship that records an initial, large erosion response in concert with the rock uplift, followed by a long period of time where the rate of erosion decreases as mean elevation and mean slope are reduced (Ahnert 1970; Pazzaglia and Brandon 1996). Impulsive increases in sediment yield over these timescales are probably correctly interpreted as some major change in the erosion processes and rates operating on a rejuvenated landscape imposed by renewed rock uplift, a change in climate, or both.

Unfortunately, use of the term rejuvenate has been extended to explain the forms and changes in individual components of a landscape over shorter timescales (10^0–10^5 yrs), but its applicability in this context is probably not correct. For example in the strict Davisian interpretation, a meandering river channel flowing in a wide river valley is a mature or even old landform whereas a steep river channel flowing in a narrow valley is a youthful landform. Individual landforms such as river channels are much better explained as a DYNAMIC EQUILIBRIUM expression between driving and resisting forces where form and process are mutually dependent. The meandering channel speaks more to the fact that the river has a stable discharge, primarily fine-grain size, gentle slopes, and stable, vegetated channel banks rather than its age in the geographic cycle. In fact, active meander channels in bedrock are known to exist in even the most rapidly uplifting landscapes such as Taiwan where there is no evidence that they have been superimposed or inherited from earlier forms (Hovius and Stark 2001; Hartshorn *et al.* 2002). Similarly, steep, narrow river valleys are common on the great ESCARPMENTs of the southern continents which are known to be among the most slowly eroding and changing landscapes on the planet (Bierman and Caffee 2001). The term rejuvenation is improperly used in these cases of attempting to explain relative landform age or changes in the landscape through an investigation of forms only, without consideration of process or tectonic setting.

References

Ahnert, F. (1970) Functional relationship between denudation, relief, and uplift in large mid-latitude drainage basins, *American Journal of Science* 268, 243–263.

Bierman, P.R. and Caffee, M. (2001) Slow rates of rock surface erosion and sediment production across the Namib desert and escarpment, southern Africa, *American Journal of Science* 301, 326–358.

Davis, W.M. (1889) The rivers and valleys of Pennsylvania, *National Geographic Magazine* 1, 183–253.

——(1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.

Hartshorn, K., Hovius, N., Dade, W.B. and Slingerland, R.L. (2002) Climate-driven bedrock incision in an active mountain belt, *Science* 297, 2,036–2,038.

Hovius, N. and Stark, C.P. (2001) Actively meandering bedrock rivers, *EOS Transactions* 82(47), 506.

Pazzaglia, F.J. and Brandon, M.T. (1996) Macrogeomorphic evolution of the post-Triassic Appalachian mountains determined by deconvolution of the offshore basin sedimentary record, *Basin Research* 8, 255–27?.

Schumm, S.A. and Rea, D.K. (1995) Sediment yield from disturbed earth systems, *Geology* 23, 391–394.

FRANK J. PAZZAGLIA

RELAXATION TIME

Geomorphological change may be envisaged as a set of responses to the varying frequencies and magnitudes of formative events at all scales (Graf 1977; Brunsden and Thornes 1979). The concept of LANDSCAPE SENSITIVITY to changes in the operation of controlling processes suggests three divisions of time (Brunsden 1980, 1990; Figure 131): the time taken to react to an impulse of change (lag or *reaction time*); the time taken to attain the characteristic state (*relaxation time*); and the time over which the form exists (*characteristic time* or *landform lifetime*) (McSaveney and Griffiths 1988). Relaxation time is an important measure because landforms can only reach a slowly changing (stable?) state if the interval between form-changing events is greater than the sum of reaction and relaxation times. If the interval is shorter then the landforms will be in a state of constant readjustment and strong flux. This state may be called *transient*.

A further application of the idea of relaxation is to define the term as '*recovery*'. After a severe or land forming event the more 'normal' frequent events will seek to erase the landform or to modify the form until it is compatible with them.

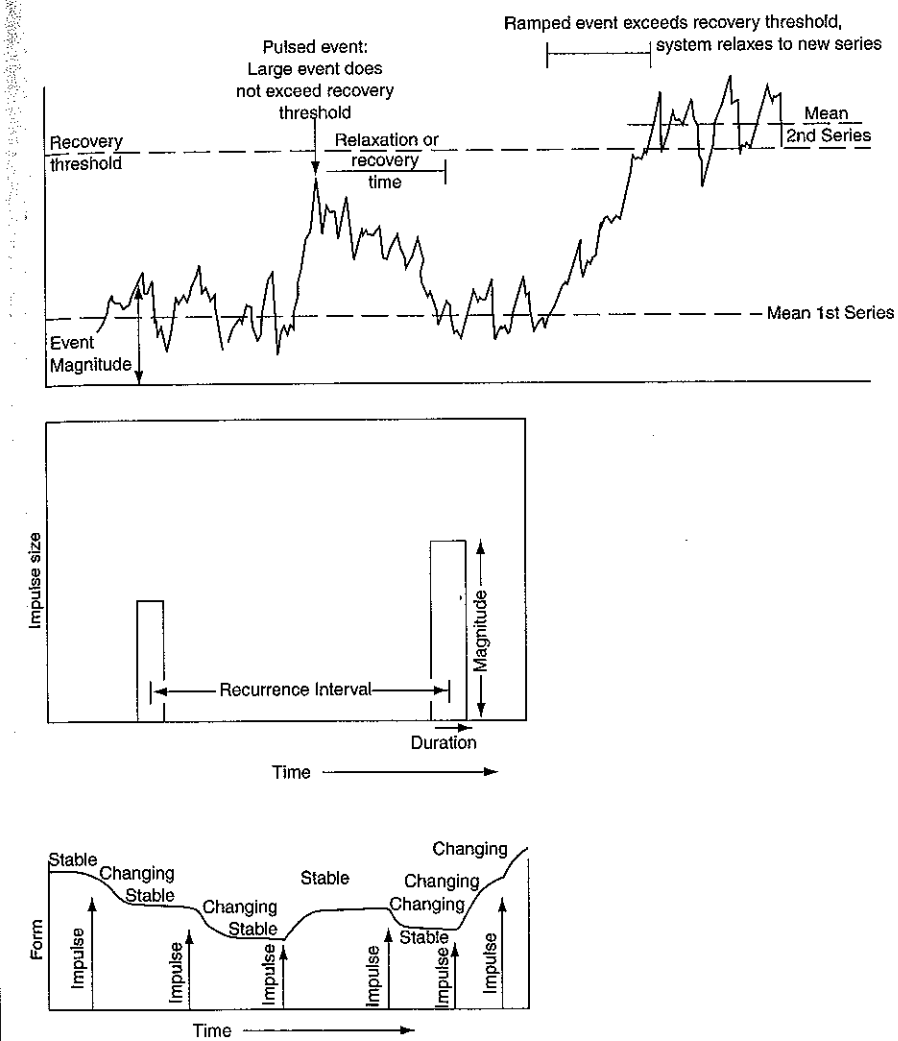


Figure 131 A schematic representation of the concepts of *reaction* (lag) time, *relaxation* (recovery, healing, form adjusting) time and *characteristic form* (form constant?) time

The process can also be an attempt to 'heal' the scars and to return the landscape to its former state. Crozier (1986; see also Crozier *et al.* 1990) regards this as a process of 'ripening' in which the landscape is again prepared for another effective event. This idea is usually applied to soil erosion and mass movement on hillslopes where hollows produced by these processes are weathered and infilled until critical depth is reached and failure can again take place (Deitrich and Dorn 1984, Deitrich *et al.* 1992).

References

Brunsden, D. (1980) Applicable models of long term landform evolution, *Zeitschrift für Geomorphologie N.F. Supplementband* 36, 16–26.

——(1990) Tablets of Stone: toward the ten commandments of geomorphology, *Zeitschrift für Geomorphologie N.F. Supplementband* 79, 1–37.

Brunsden, D. and Thornes, J.B. (1979) Landscape sensitivity and change, *Transactions Institute of British Geographers* NS4, 463–484.

Crozier, M.J. (1986) *Landslides: Causes, Consequences and Environment*, London and Dover: Croom Helm.

Crozier, M.J., Vaughan, E.E. and Tippett, J.M. (1990) Relative instability of colluvial-filled bedrock depressions, *Earth Surface Processes and Landforms* 15, 326–339.

Deitrich, W.E. and Dorn, R. (1984) Significance of thick deposits of colluvium on hillslopes, a case study involving the use of pollen analysis in the coastal mountains of southern California, *Journal of Geology* 92, 147–158.

Deitrich, W.E., Wilson, C.J., Montgomery, D.R., McKean, J. and Bauer, R. (1992) Erosion thresholds and land surface morphology, *Geology* 20, 675–679.

Graf, W.L. (1977) The rate law in fluvial geomorphology, *American Journal of Science* 277, 178–191.

McSaveney, M.J. and Griffiths, G.A. (1988) *A General Theory for Frequency Distribution of Age and Lifetime of Steepland Elements Formed by Physical Weathering*, New Zealand Geological Survey, Christchurch, NZ 1–10.

DENYS BRUNSDEN

RELIEF

Relief may be defined most generally as the elevation difference over a predetermined area or inferred length scale. This simple definition allows for specifying a number of particular types of relief. The relief of a mountain range, for example, may be considered as the difference in elevation between the highest peak and the base of the range front. Alternatively, the relief of a mountain range can refer to the absolute height of the highest peak, with the implicit reference to sea level as a datum. When defined over much shorter length scales, relief can be defined as the range in elevation spanned by a particular hillslope, ridge to valley transect, or physiographic feature such as an escarpment. Relief may also simply refer to topography in general, or more specifically to the collective elevations or their inequalities of a land surface. In other words, the term relief has a variety of possible meanings depending upon the context within which it is used. The most common use of the term, however, generally refers either to topography itself or to the elevation difference between the highest and lowest points within an area of interest.

Differences in elevation that produce relief arise from the interaction of spatial variations in rock uplift and erosion. Volcanic and tectonic processes that raise rocks above sea level are ultimately responsible for elevating mountain ranges, although normal faulting also may produce local relief in extensional settings. Erosional processes may limit the total relief maintained by rock uplift but also cut valleys and produce relief over shorter length scales. Fluvial and glacial processes that incise the landscape produce relief, whereas mass-wasting processes (such as soil creep and many types of landsliding) tend to reduce relief. The overall relief of a mountain range ultimately depends on the balance between uplift and erosion, unless accumulation of crustal material exceeds the mechanical limit supportable by crustal strength, leading to the growth of a high plateau.

Several kinds of relief can be used to describe different aspects of a drainage basin. Fluvial relief represents the elevation drop measured down the longitudinal profile of a river network, as given by the elevation difference between the channel head and the basin outlet. This portion of the total basin relief may be influenced by changes in fluvial processes and rates of river incision. Hillslope relief defined by the elevation difference from the channel head to the drainage divide at the head of the basin represents that portion of the total relief within a drainage basin beyond the immediate influence of fluvial processes and which is instead controlled by hillslope processes. The geophysical relief of a drainage basin has been described as the local elevation difference between the ridgetop and valley bottom, which consists of both hillslope and fluvial relief. In addition to these specific types of relief, local relief may be defined by the elevation difference between the highest and lowest point on the topography measured over an area of predetermined size or a proscribed length scale.

Local relief is inherently scale dependent. The larger the length scale over which it is measured, the larger the relief. Generally, local relief increases as a non-linear function of the diameter of the area over which it is measured, with an exponent <1 and typically about 0.7 to 0.8. In addition, mean local relief is strongly correlated with mean local slope. But in comparison to mean slopes, which have a strong grid-size dependence, mean local relief is less grid-size dependent when calculated from digital elevation models.

Fundamental relationships between relief and erosion rates have been posited since early workers argued that greater relief and steeper slopes lead to faster erosion. In one of the first modern studies of the influence of relief on erosion rates, Schumm (1963) reported a linear relation between erosion rate and drainage basin relief (the height above sea level of the highest point in the basin) for large North American drainage basins. Ahnert (1970) subsequently reported that erosion rates increase linearly with mean local relief (the difference in elevation measured over a specified length scale) for mid-latitude drainage basins. Later studies bolstered Ahnert's relation with data from other regions and showed that local relief and runoff are dominant controls on erosion rate for major world drainage basins (e.g. Summerfield and Hulton 1994). Different relations between erosion rates and mean elevation characterize tectonically active and inactive mountain ranges (Pinet and Souriau 1988), and Montgomery and Brandon (2002) recently reported evidence for a strongly non-linear relation between long-term erosion rates and mean local relief.

Until relatively recently, the relief of bedrock hillslopes was thought not to be strength limited because of the great cohesive strength of intact rock. But the development of discontinuities in rock strength at the scale of an entire hillslope, valley side, or mountain can limit relief development through large-scale bedrock landsliding (Schmidt and Montgomery 1995). The catastrophic 1991 failure of the crest of Mt Cook – the highest point in New Zealand – illustrates how bedrock landsliding can limit relief in steep, highly dissected terrain. Arguing for the generality of strength-limited hillslope relief, Burbank and others (1996) demonstrated that the gorge of the Indus River had strong gradients in incision rate through a region where mean hillslope gradients are independent of the local river incision rate. Hence, they concluded that the development of strength-limited hillslopes allowed bedrock landsliding to efficiently adjust slope profiles such that ridgetop lowering keeps pace with rapid bedrock river incision. This emerging view of the role of relief on erosion rates holds that in steep tectonically active regions erosion rates adjust to high rates of rock uplift primarily through changes in the frequency of landsliding rather than increased hillslope steepness or increased relief (Montgomery and Brandon 2002). In contrast, in lower gradient landscapes the steepness of hillslopes, and therefore local relief, may respond to changes in the controls on landscape-scale erosion rates.

Climate setting and variability constrain the total relief of mountain ranges. Highly orographic rainfall variability can either limit or increase the fluvial relief depending upon the nature of the feedback operating in a specific mountain range (Roe *et al.* 2002). Enhanced erosion by glaciers and periglacial processes can preclude development of relief substantially above the perennial snowline (Brozovic *et al.* 1997). The role of erosion in reducing mass accumulation in mountain ranges is perhaps best illustrated by the exceptional cases where lack of rainfall allows mass accumulation to engage the mechanical limit to crustal thickening and results in development of high plateaux like the Altiplano and Tibet. The position of Earth's high plateaux in the dry latitudes suggests that plateau formation reflects the coincidence of high rates of tectonically driven mass convergence and low rates of erosion due to an arid climate.

A new view of the coupling and feedback among climate, erosion and tectonic processes is coalescing from recent studies focused on their interactions. Geologists are recognizing that spatial gradients in the climate forcing that drives erosion can influence the development and evolution of geologic structures. Development of mountain ranges strongly influences patterns of precipitation and numerical simulations of evolving and steady-state orogens show that both the topography and the resulting metamorphic gradients exposed at the surface reflect the influence of spatial variability in erosion (Willett 1999). Gradients in climate and tectonic forcing strongly influence erosional intensity, and this interaction in turn governs the development and evolution of topography. Hence, the development of relief is strongly coupled to large-scale feedback involving the interplay of climate, erosion and tectonics. Whereas the geographical distribution of plate tectonic environments has changed over geologic time, the global pattern of climate variability exhibits robust latitudinal patterns characterized by abundant and intense rainfall in the equatorial tropics, a low latitude belt of deserts and stronger glacial influences toward the poles. In this context, the feedback between climate, tectonics and erosion implies large-scale climatic controls on the global distribution of topography; high plateaux are likely to form astride the desert latitudes, whereas high mountains are unlikely to form in the equatorial or polar regions where erosion rates are high due to either intense rainfall or glacial processes.

Substantial debate has centred on the relation of climate change to the relief of mountainous topography. An increase in the absolute relief of a mountain range, and consequent increase in the area of alpine environments, can increase rates of weathering through rapid mechanical breakdown of fresh rock by periglacial and glacial processes. Hence, increased relief in alpine areas potentially could influence global carbon cycles and large-scale climate. Wager (1933) noted the proximity of high Himalayan peaks to deep valleys and proposed that isostatic rebound in response to valley incision was responsible for elevating Himalayan peaks above the Tibetan Plateau. Molnar and England (1990) proposed that much of the evidence for substantial late Cenozoic uplift of mountain ranges may simply represent the effect of climatic deterioration on increased erosional exhumation of rocks or the uplift of mountain peaks in response to deepening and enlargement of valleys. Analyses of valley geometry show that such an effect could account for up to about a quarter of the elevation of mountain peaks, although the potential magnitude of such an effect depends on the strength of the crust and the nature of erosional processes (Gilchrist *et al.* 1994; Montgomery 1994). However, recent studies have concluded that there is minimal potential for valley incision to substantially influence local relief in tectonically active mountain ranges (Whipple *et al.* 1999; Montgomery and Brandon 2002).

In summary, relief is a simple concept with many variants of meaning that depend on the specific context in which it is used. Nonetheless, an understanding of the controls on relief generation is central to understanding the linkages between geomorphic processes, tectonics and climate that together shape the Earth's surface.

References

Ahnert, F. (1970) Functional relationship between denudation, relief, and uplift in large mid-latitude drainage basins, *American Journal of Science* 268, 243–263.

Brozovic, N., Burbank, D.W. and Meigs, A.J. (1997) Climatic limits on landscape development in the Northwestern Himalaya, *Science* 276, 571–574.

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.

Gilchrist, A.R., Summerfield, M.A. and Cockburn, H.A.P. (1994) Landscape dissection, isostatic uplift, and the morphologic development of orogens, *Geology* 22, 963–966.

Molnar, P. and England, P. (1990) Late Cenozoic uplift of mountain ranges and global climate change: chicken or egg? *Nature* 346, 29–34.

Montgomery, D.R. (1994) Valley incision and the uplift of mountain peaks, *Journal of Geophysical Research* 99, 13,913–13,921.

Montgomery, D.R. and Brandon, M.T. (2002) Non-linear controls on erosion rates in tectonically active mountain ranges, *Earth and Planetary Science Letters* 201, 481–489.

Pinet, P. and Souriau, M. (1988) Continental erosion and large-scale relief, *Tectonics* 7, 563–582.

Roe, G.H., Montgomery, D.R. and Hallet, B. (2002) Effects of orographic precipitation variations on the concavity of steady-state river profiles, *Geology* 30, 143–146.

Schmidt, K.M. and Montgomery, D.R. (1995) Limits to relief, *Science* 270, 617–620.

Schumm, S.A. (1963) *The Disparity between Present-day Denudation and Orogeny*, US Geological Survey Professional Paper 454-H.

Summerfield, M.A. and Hulton, N.J. (1994) Natural controls of fluvial denudation rates in major world drainage basins, *Journal of Geophysical Research* 99, 13,871–13,883.

Wager, L.R. (1933) The rise of the Himalaya, *Nature* 132, 28.

Whipple, K.X., Kirby, E. and Brocklehurst, S.H. (1999) Geomorphic limits to climate-induced increases in topographic relief, *Nature* 401, 39–43.

Willett, S.D. (1999) Orogeny and orography: the effects of erosion on the structure of mountain belts, *Journal of Geophysical Research* 104, 28,957–28,981.

DAVID R. MONTGOMERY

RELIEF GENERATION

Almost everywhere landforms are composed of elements evolved under different climates, i.e. shaped by different exogenetic forces. The relevant form assemblage is called relief generation. They are the constituents of CLIMATO-GENETIC GEOMORPHOLOGY. This concept has no relation to the stages of the Davisian CYCLE OF EROSION and its DENUDATION CHRONOLOGY, which are based on tectonics.

In central Europe, rumpfflächen (etchplains, plains cutting rocks of different hardness; see ETCHING, ETCHPLAIN AND ETCHPLANATION) came

into existence in Tertiary time as is concluded from the form, deposits and relics of tropical weathering. Into these plains valleys are incised, starting with broad terraces of Pliocene/lower Pleistocene age with almost pure quartz gravels, sometimes with a few pisoliths. Obviously they are the eroded and transported result of tropical weathering. The flight of terraces in the middle and lower part of the valleys is of periglacial origin as is proved by pebbles from different rocks, syngenetic permafrost features, and LOESS or dunes (only Würm) on top. A solifluction cover on almost every slope, as strata in which the recent soils are developed, shows the overall small amount of Holocene erosion. This applies too for fluvial processes as the floodplain is only about 3 m below the Würm terrace and the incision was mainly in the late Würm, or early Holocene. The term relief generations was introduced by Büdel in 1955 in a paper about the Hoggar in the central Sahara. Here the rumpfflächen carry red loam, relics of tropical soils, under dated basalt flows. A loams terrace in the valleys does not correspond with the recent processes, and has very old artefacts on top. The recent river bed consists of sand.

These examples show the main methods to distinguish relief generations: (1) separation of landforms of different origin as the younger ones are nested or incised into the older; rarely the younger form is on top of the older ones as for instance dunes; (2) observation of the recent processes thus delineating the recent forms; (3) search for weathering relics and/or correlated sediments giving an indication of a different climate and linking these to the older relief generation. If a complete picture is derived by observations in the field, perhaps added to by laboratory analysis, one then tries to compare the forms of the relief generation with similar forms in a different climatic zone. Absolute datings are helpful as they provide an age, which, via the geologic timescale, gives an idea of paleoecological conditions.

The basis for the comparison is 'Klimatische Geomorphologie', which is quite different from CLIMATIC GEOMORPHOLOGY, which studies the landform assemblage and the relative importance of the recent processes in morphoclimatic zones. A regular assemblage of landforms presents the chance to classify relic forms that are only sporadically preserved and to search for additional forms. For example, if one observes overdeepening and glacial striations in rocks, a wall of mixed deposits in a certain position, then it is most likely a moraine. This can be backed up by looking for drumlins or other features of glacial erosion. The concept of relief generations is broader than investigating palaeoforms, for it asks for form assemblages and their relief forming mechanisms. After all these investigations one might ask for palaeoclimatic data from e.g. palaeobotany for comparison. It has almost never been tried to fix recent climatic data to the boundaries of regions with different relief generations.

The relief forming processes of different climates in Tertiary to recent times might also be seen at the small scale. Blocks in blockfields in the Harz Mts have a red rind from weathering in the Tertiary red loam. This is topped by a small white rind. On several blocks a triangular or square piece has been split off by frost weathering. Here the edges of the block have a white rind only. On the rim of the blockfield further blocks are uncovered by recent wash as most probably happened during warm periods of the Pleistocene, too. Thus these fields may be called 'Mehrzeitformen' (multitude forms of different climates). Ayers Rock has weathering forms in the hard rock, which are nested and which were formed in different climates.

For methodological discussions relief generations should be the basis for the distinction of relief elements of different sensitivity to recent geomorphological processes. Elements of older relief generations are more stable than younger ones, and they may be eroded mainly by valley incision. The strength and place of occurence of thresholds can be explained by relief generations, e.g. the edge of an old plain. The elements of older relief generations are certainly not in equilibrium with recent processes. They have not been formed by them nor are they considerably changed by them. Equilibrium is almost never provable even for the recent generation for two reasons: the influence of older generations on the discharge and water movement paths, and the different resistance of existing forms. The ergodic principle can only be applied to relief forms of one generation, preferably those which change fast. Thus the ergodic principle is limited in time.

Further reading

Büdel, J. (1977) *Klima-Geomorphologie*, Berlin. Translated by L. Fischer and D. Busche (1982) *Climatic Geomorphology*, Princeton: Princeton University Press.

Bremer, H. (1965) Ayers Rock, ein Beispiel für klimagenetische Morphologie, *Zeitschrift für Geomorphologie N.F. Supplementband* 9, 249–284.

HANNA BREMER

REMOTE SENSING IN GEOMORPHOLOGY

Remote sensing is the acquisition of information about an object without physical contact. In geomorphology, remote sensing often implies the collection of information from aerial platforms (e.g. airplanes, balloons or kites) or from spacecraft orbiting the Earth. The term remote sensing is credited to Evelyn L. Pruitt and her staff in the United States Office of Naval Research. It was coined during the early 1960s in recognition that instruments other than cameras and regions of the electromagnetic spectrum outside those visible to the human eye and to which photographic film is sensitive were increasingly being used to image the Earth. The current American Society of Photogrammetry and Remote Sensing's (ASPRS) definition of photogrammetry and remote sensing reads 'the art, science, and technology of obtaining reliable information about physical objects and the environment through the process of recording, measuring, and interpreting imagery and digital representations of energy patterns derived from noncontact sensor systems' (Colwell 1997: 3).

While remote sensing will not replace the traditional geomorphic field study, the value of remote sensing to provide a synoptic overview of a landscape cannot be overlooked. Historically, the use of remote sensing in geomorphology has been mainly interpretive, enabling geomorphologists to develop a 'mental picture' of the landscape and as a map-making aid (Hayden *et al.* 1986). However, the use of remote sensing for quantitative geomorphic study is growing rapidly.

Remote sensing provides unique global views at different spatial scales and in different regions of the electromagnetic spectrum. These global views are extremely useful for the subdiscipline of megageomorphology, which emphasizes the study of planetary surfaces at large scales (Baker 1986). The global views provided by remote sensing are not static, but are being continuously refreshed. This repeated global monitoring captures geomorphic events that might otherwise go unnoticed. For example, in 1983 astronauts aboard the STS-8 Space Shuttle mission photographed a dust storm over northwestern Argentina that was transporting material from exposed salt flats on the Puna Plateau of the South American Andes eastward toward the Argentine Pampas. These remote observations helped confirm the source of the Pampas loess and demonstrated that silt accumulation on the Pampas is an ongoing geomorphic process (Hayden *et al.* 1986). Remote sensing can also enable geomorphic study of areas that are inaccessible to field-based investigations.

Remote sensing provides a unique historical archive of geomorphic change. Aerial photography suitable for geomorphic analysis began to be collected as early as the 1920s. The satellite image archive suitable for geomorphic analysis began with the launch of the first of the Earth Resources Technology Satellite satellites ERTS-1, later renamed to Landsat-1, on 23 July 1972.

The history of remote sensing as a tool for geomorphic analysis is intimately tied to advances in photography and the acquisition of photographs from aerial platforms. Photography was born in 1839 when the photographic processes developed by Joseph Nicephore Niepce, Louis Jacques Mande Daguerre and William Henry Fox Talbot were publicly disclosed. One year later the use of photography to aid in the development of topographic maps was advocated by François Arago, Director of the Paris Observatory (Fischer 1975: 27). The first aerial photograph was taken by Gaspard Felix Tournachon, also known as Nadar, from a balloon outside Paris, France in 1858. However, recognition of photography's value to geomorphology also arose in part from terrestrial photographs of the landscape taken during the latter half of the nineteenth century. In 1890, the Geological Society of America formed a Committee on Photographs and its first report described photogeology (Fischer 1975: 34) which as a science took shape in the 1920s and 1930s. The basis of photogeologic analysis rests on the simple notion that landforms developed under similar geologic and geomorphic processes will appear similar in remotely sensed images (Way and Everett 1997: 117). By the early 1940s, photo interpreters were able to recognize the distinguishing surface features of approximately thirty-five major landforms and realized that aerial photographs provided important information on the origin, composition and history of landforms (Colwell 1997: 26). These landforms, and other features, can be identified in remote sensing images based on their location, size, shape, tone and/or colour, shadow, texture, pattern, height/depth as well as site characteristics and associations among features in the landscape (Jenson 2000: 121–132). The first systematic, although low resolution, observation of Earth from satellite began in 1960 from TIROS I, the world's first meteorological satellite. Since then, remote sensing of the Earth has expanded significantly as spaceborne imaging systems have grown in number and in sophistication.

Successful application of remote sensing images to geomorphic study requires careful matching of an instrument or image's spatial, temporal and spectral characteristics with the requirements of the geomorphic study at hand. A sensor's spatial resolution is the smallest angular or linear separation that it can resolve. The resolution of digital images acquired by non-photographic instruments is often described by the length of one dimension (in metres) of the individual elements (pixels) that comprise the two-dimensional image. In determining the required spatial resolution required for a particular application, a useful rule of thumb is that for an instrument to detect a feature of a certain size, its spatial resolution should be at most one-half of the feature's smallest dimension.

The temporal resolution of a remote sensing system refers to how frequently it can image a certain area. Most orbital sensors have fixed repeat cycles which control how often an area is imaged. They typically range from less than one day to one or two weeks. Aircraft overflights or manned space missions usually acquire images much more infrequently and at much more irregular intervals.

Various wavelength regions of the electromagnetic spectrum provide quite different information about the chemical, physical and biological properties of a landscape. The two most commonly used regions of the electromagnetic spectrum for geomorphic study are the optical, where the propagating electromagnetic energy has wavelengths of 0.3 to 14 micrometres (μm, 1 μm = 10^{-6} m), and the microwave with millimetre to metre wavelengths.

The optical region, which historically has been the most widely used, can be divided into two subregions, a reflected optical region (0.3–3.0 μm) and a thermal infrared region (3.0–14.0 μm). Remote sensing instruments operating in these two wavelength regions are typically passive; the energy supplying the signal to the sensor comes from an external source. In the reflected optical region, solar energy reflected off the landscape provides the signal while in the thermal infrared region, energy emitted directly by the Earth itself as a function of its surface temperature is the primary energy source. These two spectral regions contain numerous atmospheric windows or wavelength intervals in which the atmosphere is fairly transparent to solar energy or emitted energy making remote observations possible. While the atmosphere may be fairly transparent in these windows, for some geomorphic applications correction for atmospheric effects still may be important. Cloud cover can also obscure the surface over all wavelengths in the optical.

The reflected optical wavelengths are the most commonly used in terrestrial remote sensing. The reflected optical is typically subdivided into three spectral regions: the visible (0.4–0.7 μm) to which the human eye is sensitive, the near infrared (0.7–1.1 μm) and mid or short-wave infrared (1.1–2.5 μm). As the name suggests, reflected optical images are formed from the energy reflected from the surface towards the sensor. The reflectance of surface materials varies as a function of wavelength making it often possible to discriminate between different surface materials based solely on their reflectance. While two materials may appear similar at one wavelength they may be quite easy to distinguish at another.

Remote sensing instruments can also provide a valuable three-dimensional view of a landscape through the use of stereopairs, which are two remote sensing images that when viewed together add the illusion of relief to a landscape. Stereoscopic measurements can be used to make topographic maps or digital representations of topography known as digital elevation models (DEMs). Stereopairs also are stellar in their support of one of the original stated purposes of photogeology which was 'to provide better illustrations for teaching geology' (Fischer 1975: 34). Two excellent modern examples of the educational value of stereopairs and satellite images for teaching about the Earth's landforms are the *Atlas of Landforms* (Curran *et al.* 1984) and *Geomorphology from Space* (Short and Blair 1986).

The microwave region (wavelengths ranging from one mm to one m) of the electromagnetic spectrum is also important for geomorphic remote sensing. Synthetic Aperture Radars (SARs) are active remote sensing instruments that illuminate the ground with their own electromagnetic signal and then record the amount of energy that is scattered from the target back to the sending antenna. Therefore, SAR images are sometimes referred to as backscatter images. SARs offer advantages over optical sensors as they can penetrate through clouds and obtain images at night making them ideal for studying cloudy regions and for capturing short-lived dynamical geomorphic processes like flooding. SAR images capture quite different characteristics of the landscape than do optical sensors. The backscatter signal received at a SAR antenna is affected by the roughness of the surface and the moisture present in the soil and in vegetation. This makes SAR useful for assessing such important landscape properties as soil moisture, melting conditions on the surface of glaciers, the biomass of plant communities and flooding or inundation.

One unique feature of SAR that has proved valuable for geomorphic research is the ability of SAR to penetrate into dry materials, such as sand or dry snow. The longer the SAR wavelength the deeper into the subsurface the microwave energy can penetrate. A classic geomorphic study demonstrating SAR's ability to provide subsurface information was the identification of an extensive drainage network under the Selima Sand Sheet covering portions of western Egypt and eastern Sudan not easily visible from field observation or optical sensors (Hayden *et al.* 1986).

In geomorphology, remote sensing is not limited to the collection of images of terrestrial surfaces from the air or space. Ground-based remote sensing techniques including ground penetrating radar (GPR) and seismic reflection profiling provide detailed two or three-dimensional images of the near subsurface useful for studying the internal structure of landforms and glaciers and for environmental site analysis. Since the 1960s, multibeam acoustical sounding instruments, often known as side scanning SONARs (sound, navigation and ranging), have been widely used in marine geomorphology and bathymetric mapping. Remotely operated vehicles (ROVs) are providing fascinating views of otherwise unseen marine environments. Remote sensing of landscapes is not limited to Earth. By necessity, planetary geologists have made extensive use of remote sensing to study other planets and even asteroids in our solar system (Hayden *et al.* 1986).

References

Baker, V.R. (1986) Introduction: regional landform analysis, in N.M. Short Sr and R. Blair Jr (eds) *Geomorphology from Space, NASA SP-486*, Washington, DC: National Aeronautics and Space Administration.

Colwell, R.N. (1997) History and place of photographic interpretation, in W.R. Philipson (ed.) *Manual of Photographic Interpretation*, 2nd edition, 3–47, Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

Curran, H.A., Justus, P.S., Young, D.M. and Garver, J.B. (1984) *Atlas of Landforms*, 3rd edition, New York: Wiley.

Fischer, W.A. (1975) History of remote sensing, in R.G. Reeves (ed.) *Manual of Remote Sensing*, 1st edition, 27–50, Falls Church, MD: American Society of Photogrammetry.

Hayden, R.S., Blair, R.W. Jr, Garvin, J. and Short, N.M. Sr (1986) Future outlook, in N.M. Short Sr and R. Blair Jr (eds) *Geomorphology from Space*, NASA SP-486, Washington, DC: National Aeronautics and Space Administration.

Jenson, J.R. (2000) *Remote Sensing of the Environment: An Earth Resource Perspective*, Upper Saddle River, NJ: Prentice Hall.

Short, N.M. Sr and Blair, R.W. Jr (1986) *Geomorphology from Space. A Global Overview of Regional Landforms*, NASA SP-486, Washington, DC: National Aeronautics and Space Administration. Available online at: http://daac.gsfc.nasa.gov/DAAC_DOCS/geomorphology/GEO_HOME_PAGE.html

Way, D.S. and Everett, J.R. (1997) Landforms and geology, in W.R. Philipson (ed.) *Manual of Photographic Interpretation*, 2nd edition, 117–165, Bethesda, MD: American Society for Photogrammetry and Remote Sensing.

ANDREW KLEIN

REPOSE, ANGLE OF

The maximum angle at which a mass of debris under given conditions will remain stable. The angle of repose generally varies between 25° and 40°. For instance, the angle of repose for sand is between 30° and 35°, whereas for scree it is between 32° and 36°. The exact angle of repose depends upon slope conditions such as the size, shape, roughness and degree of interlocking, sorting, the height of fall, and density of the individual sediment grains. Also, the length of slope and the pore-water pressure of the sediment are important, as increased water content enhances structural integrity of the sediment due to surface tension between grains. A general understanding of the angle of repose is known, though studies concerning factors influencing the angle of repose have produced diverse results.

Further reading

Francis, S.C. (1986) The limitations and interpretation of the 'angle of repose' in terms of soil mechanics; a useful parameter? in A.B. Hawkins (ed.) *Engineering Geology Special Publication* 2, 235–240.

Kinya, M., Kenichi, M. and Shosuke, T. (1997) Method of measurement for the angle of repose of sands, *Soils and Foundations* 37(2), 89–96.

STEVE WARD

RESIDUAL STRENGTH

Also termed ultimate strength. It refers to the minimum remaining degree of strength (i.e. resistance to movement) in a soil or rock after loss of strength following significant displacement (relative to the material but typically >1 m). The term is thus linked with slope movements, and is extremely important in slope stability analysis in order to gauge the strength of a pre-existing active slope. Residual strength in sands is typically the same as the critical shear strength (a steady state subsequent to shearing in which the effective stresses remain constant and no volume changes occur), whereas materials with high levels of clay provide a residual value of about half the critical shear strength. Soils high in platy clay materials cause a considerable reduction in strength (from peak to residual) as they tend to align themselves parallel to the direction of displacement following movement.

Further reading

Carrubba, P. and Moraci, N. (1993) Residual strength parameters from a slope instability, in P. Shamsger (ed.) *Third International Conference on Case Histories of Geotechnical Engineering*, 1,481–1,486, Rolla: University of Missouri.

Spangler, M.G. and Handy, R.L. (1982) *Soil Engineering*, New York: Harper and Row.

STEVE WARD

REYNOLDS NUMBER

A dimensionless number used in fluid dynamics to determine the transition from laminar to turbulent flow through a pipe, developed by Osborne Reynolds in 1883. The parameter is based upon the fact that the ratio of kinetic energy to energy transferred by viscous forces was correlated with turbulent flow. The number is defined by the equation $Re = V\,L / v$, where Re is the Reynolds number, V is the velocity, L is the length, and v is the kinematic viscosity (viscosity/density). When this ratio is less that 1,000 laminar flow will be observed, whereas high Reynolds numbers represent turbulent flow. However, the actual definition of a high Reynolds number is determined by the shape of the system. Reynolds also studied the effects of flow resistance in pipes, demonstrating that the friction coefficient is a unique function of the Reynolds number at various surface roughnesses.

Further reading

Reynolds, O. (1883) An experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous and of the law of resistance in parallel channel, *Philosophical Transactions of the Royal Society* 174, 935–982.

Rott, N. (1990) Note on the history of the Reynolds number, *Annual Review of Fluid Mechanics* 22, 1–11.

SEE ALSO: boundary layer

STEVE WARD

RIA

A coastal inlet resulting from the drowning of a former river valley system or estuary. Rias are formations originating during the postglacial glacioeustatic transgression of the seas across the continental shelf during the Flandrian TRANSGRESSION, following the melting of the ice sheets and glaciers. This resulted in the development of an extremely irregular, indented coastline, where only the pre-existing hill peaks remained above sea level. Rias are non-glaciated, having been formed originally by subaerial erosion, and are characteristically long, narrow, often funnel-shaped inlets, whose depth and width uniformly decreases inland. They are also shorter and shallower than a fjord. The term ria originates from the type locations of Galicia and Asturias, north-west Spain, where a series of long mountainous-sided estuaries exists, once drowned by postglacial eustatic sea-level rise. Other examples include the south-west of Ireland (Kerry and Bantry Bay). A less restricted use of the term ria exists, pertaining to any broad estuarine river mouth, including fjords. However, the original application of the term is preferred in geomorphology.

Further reading

Cotton, C. (1956) Rias *sensu stricto* and *sensu lato*, *Geographical Journal* 122, 360–364.

STEVE WARD

RICHTER DENUDATION SLOPE

A hillslope type, common in extreme environments such as alpine and polar regions, that develops through cliff retreat and forming a uniform (rectilinear) slope at the angle of rest of the accumulating talus. Richter denudation slopes were first noted by E. Richter in 1900, following studies in the Alps. The formation of such slopes was later expanded on by Bakker and Le Heux (1952), who modelled Richter slope formation. Richter denudation slopes develop essentially by rock fall, where the resulting talus is moved gradually by rolling and sliding, and forming a thin veneer over the basal slope. The cliff retreats steadily, often cutting across bedrock, while the basal slope may either accumulate talus, thus raising the foot of the slope, or be removed by weathering or abrasion of sliding talus. The free face will eventually be eliminated resulting in a smooth hillslope of uniform gradient. Examples of such slopes are common in the Transantarctic Mountians and Koettlitz Valley, Antarctica.

Reference

Bakker, J.P. and Le Heux, J.W.N. (1952) A remarkable new geomorpholgical law, *Koninklijke Nederlandsche Akademie von Wetenschappen* B55, 399–410, 554–571.

Further reading

Selby, M.J. (1982) *Hillslope Materials and Processes*, Oxford: Oxford University Press.

STEVE WARD

RIDGE AND RUNNEL TOPOGRAPHY

Ridge and runnel topography comprises a series of alternating intertidal bars and troughs and is typically found on sandy beaches in fetch-limited, macrotidal coastal environments. The number of ridges (and runnels) is 3–6, the height of the ridges ranges from 0.5 to 1 m and the distance between the ridges varies from 50 to 100 m. The intertidal gradient of ridge and runnel beaches is approximately 0.015, but the seaward slope of the ridges is significantly steeper and may be up to 0.05. Storm wave conditions result in a flattening, or even destruction, of the ridge morphology. Calm wave conditions, on the other hand, induce ridge build-up and promote onshore migration of the ridges. Ridge and runnel topography is relatively stable and rates of onshore bar migration rarely exceed 1 m per tide.

It was previously thought that the ridges develop as swash bars during stationary tide conditions. However, some doubt has been cast on a swash origin of the ridges and it seems more likely that the ridges are breaker bars. Whatever its origin, ridge and runnel topography is subjected to a range of hydrodynamic processes over a tidal cycle, including swash, surf and shoaling wave processes. Depending on the wave/tide conditions and the position on the beach profile, the ridges will be affected and controlled to varying degrees by each of these hydrodynamic processes.

In the American coastal literature, welded bar systems that develop following storm erosion (see BEACH) are also sometimes referred to as ridges and runnels. This usage of the term 'ridge and runnel topography' is considered inappropriate and should be avoided.

Further reading

Orford, J.D. and Wright, P. (1978) What's in a name? – descriptive or genetic implications of 'ridge and runnel' topography, *Marine Geology* 28, M1–M8.

GERHARD MASSELINK

RIEDEL SHEAR

Refers to a conjugate set of overlapping *en echelon* faults which develops during the early stages of shearing, usually at inclinations of ~15° (R Shears or synthetic fractures) and ~75° (R' Shears or antithetic fractures) to the principal displacement zone (PDZ) boundary. Riedel shear zones form in dip-slip fault regimes and are composed of fault and fracture elements marked by standard physical properties of brittle shear zones (e.g. slickenside surfaces, slickenlines, gouge and/or breccia and abundant fracturing), though others can be observed as deformation bands and zones of deformation (Davis *et al.* 2000). Second *en echelon* synthetic and antithetic shears, termed P-shears, may form through development of the Riedel shear system, although P-shears can sometimes develop before R-shears or at the same time. Lesser shears that can develop in relation to Riedel shear include Y-shears, and T fractures. Riedel shearing was first observed by Cloos (1928) and Riedel (1929) during studies on clay-cake deformation.

References

Cloos, H. (1928) Experimenten zur inneren Tektonic, *Centralblatt für Mineralogie, Geologie und Paleontologie* 1928B, 609.
Davis, G.H., Bump, A.P., Garcia, P.E. and Ahlgren, S.G. (2000) Conjugate Riedel deformation band shear zones, *Journal of Structural Geology* 22(2), 169–190.
edel, W. (1929) Zur mechanic geologischer bruderscheinungen, *Centralblatt für Mineralogie, Geologie, und Paleontologie* 1929B, 354.

E ALSO: shear and shear surface

STEVE WARD

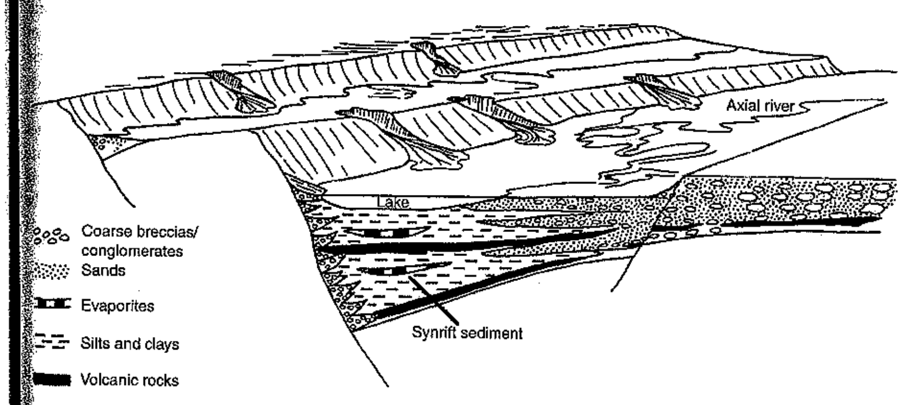
IFT VALLEY AND RIFTING

ift valleys are elongate depressions in the Earth's rface that are formed as a result of extension in e crust and upper mantle. Many are large ough to be easily visible from space from where ey resemble large cracks that cut across continental landmasses (Plate 95). At ground level they e well defined and easily recognizable geomorphological features that have been the subject for xploration and research for more than a century. The term 'rift valley' was first used by regory in the nineteenth century to describe the ast African Rift (alternatively called the Great ift or the Gregory Rift) which runs from Afar in e north of Ethiopia to Blantyre to the south of ake Malawi – a total length of 35,000 km. In thiopia this rift reaches its greatest depth of over km. Other well-known rifts include the Baikal ift of south-central Siberia, the Rhine Graben of urope and the Rio Grande Rift of western USA.

The morphology of rifts is always similar with central depression or rift valley flanked on both sides by uplifted areas. The uplifted shoulders of rifts are each associated with a staircase of (mostly) normal faults of varying magnitude which step the underlying rocks down towards the central depression.

The classical view of the structure of a rift was that it is symmetrical, with a flat bottom, flanked by two equally large border faults, a structure for which the German word 'graben' was coined. This interpretation was based almost entirely on an assumption that the surface expression of a rift is the same as the structure, ignoring the importance of erosion and deposition in modifying the landscape. The flat bottom observed e.g. in the East African Rift is due to deposition of lake and river sediments (Frostick and Reid 1989). Seismic evidence from rifts worldwide has

Plate 95 False colour satellite image of part of the Kenyan section of the East African Rift. Rift valley and other sediments show white



igure 132 Schematic diagram of a cross section through rift structure. Note the asymmetry with a hicker package of sediments filling the basin close to the border fault

shown that the dominant structure is an asymmetrical half graben with one margin more intensely faulted than the other (Figure 132). The location and character of the main border faults is controlled by the structure and lines of weakness in the pre-rift rocks. In some areas there are a few large faults, with vertical displacement in excess of 2 km, and in others a plethora of smaller ones. The rocks between the faults are tilted away from the rift axis forming parallel valleys between tilted fault blocks. Another characteristic of the faulted margin is uplift, as the valley floor is displaced downwards the shoulder of the rift rises upwards to emphasize the topographic step. Rift valleys vary in width from less than 30 to over 200 km. Along the less faulted margin there are smaller faults inclined both towards and away from the rift axis (antithetic and synthetic faults). Most of these are also normal extensional faults that carve the pre-rift rocks into a series of small HORST blocks with intervening valleys.

Although continental rifts can be viewed overall as continuous elongate depressions, it is interesting to note that the underlying structure, and to some extent its topographic expression, is segmented into a series of smaller basins which vary in length from tens to hundreds of kilometres. At the divides between basins there are transverse or oblique structural elements, the nature of which is the matter of debate, named variously transfer zones, accommodation zones, relay ramps, relay zones and segment boundaries. Across these zones the basin margin occupied by the major fault can alter, giving a sinuous form to the deepest zone of the valley along rift axis.

The origin of rifts has been the subject of great debate over many decades. Rifts develop in a variety of plate tectonic (see PLATE TECTONICS) settings and can be formed anywhere the crust of the Earth is placed under tension. The most obvious circumstance in which this will occur is when a continent is splitting apart to form a new ocean (e.g. in the Red Sea–Gulf of Aden; see Girdler 1991), but it can also occur where plates are moving laterally past or even towards each other in a non-uniform way that places local areas of the crust under tension (e.g. the Dead Sea pull-apart basin). Some researchers favour classifying rifts into those associated with the constructive plate margins that lead to the development of oceans and those that are within a plate that is not splitting apart, so called intraplate settings. However, this division is difficult to justify given that continental rifts may be in an intraplate setting but still associated with oceanic development. This is the case with both the Benue trough in west Africa, which was formed during the opening of the Atlantic Ocean, and the East African Rift which is associated with the opening of the Red Sea–Gulf of Aden (Frostick 1997). Both are failed arms of oceans that had the potential to become mid-ocean ridges. Such failed oceanic rifts are called aulacogens.

The biggest rift systems in the world are on the ocean floor in the centres of mid-ocean ridges. The worldwide network of ocean ridges constitutes the most significant topographic feature on the Earth's surface, surpassing even the Himalayas in scale. A typical ridge is 1,000–2,000 km wide and 2–3 km high. The central rift is the focus of intense earthquake activity and volcanicity.

Although it is well known that rifts form as a consequence of crustal extension, the cause of the instability has been hotly debated. The cause might be convection cells in the mantle that pull apart areas of the crust or the crust might be placed under tension by other plate movements. Whatever the mechanism, the stretching of the Earth's crust to form a rift causes it to thin in a manner similar to the thinning of semi-solid toffee as it is pulled apart. Hot, low-density mantle material wells up close to the surface resulting in high heat flows in and around most rifts. The topographic expressions of hot material from lower in the Earth penetrating closer to the surface can be the development of large domes and extensive volcanic activity. The development of large uplifted domes is a feature of the early stages of ocean opening. For example, examination of the topography and drainage of the West African margin reveals a series of large domes approximately 1,000 km in diameter that predate the opening of the Atlantic Ocean (Summerfield 1991). Similar structures are associated with the East African Rift, centred on Robit in Ethiopia and Nakuru in Kenya.

Rifts are often the focus for volcanic activity which commences at an early stage of rift development and can be extensive, for example in East Africa where an area of over 500,000 km^2 is covered with rift-related volcanic rocks and many of the well-known mountains are volcanoes including Ol Doinyo Lengai, Kilimanjaro and Mount Kenya. The nature of the volcanic rocks in rifts is distinctive and contains high concentrations of so-called volatile elements (particularly carbon dioxide and halogens). Rock types include basalts, trachytes, tuffs and carbonatites. Salts leached from these rocks can contribute to the development of saline lakes, e.g. Lakes Natron and Magadi.

The new topography that results from the development of a rift valley in a continental landmass will impact on hydrology, climate and ecology in a variety of different ways. Uplift along the rift margins reduces the ambient temperature and tends to increase rainfall while the centre of the rift remains warmer and can be more arid, depending on the latitude of the rift. Rift flanks are often the sites of more lush and temperate vegetation which, in the tropics, can form rainforest. Both the topography and the contrast in habitat from flank to valley bottom act as barriers to the migration of animals and, to a lesser extent, plants. The relatively isolated environment of the rift valley bottom is one that has played a unique role in human evolution. It is now widely accepted that the hominids found in the East African Rift, largely by members of the Leakey family, show that critical stages in the evolution of hominids occured in this area prior to migration out of Africa.

The evolution of rift morphology will disrupt the pre-existing continental drainage patterns, reversing, diverting and beheading river systems in a systematic and effective way. Pre-rift, most continental drainage systems comprise a limited number of very large, long-lived rivers fed by a well-integrated network of smaller streams and draining towards the nearest ocean margin. The impact of the incipient rift will depend upon the orientation of the new structure to the existing river system. If the rift is aligned with the main drainage direction it might capture all or some of a local river system. In contrast, a rift which cuts across the pre-existing drainage often diverts and reverses sections of the drainage. Domed sections of a rift are particularly effective at drainage diversion and develop a radial stream pattern that diverts all but very local and small rivers away from the rift basin. As the structure develops further, and faults begin to carve the surface into a series of ridges, there are new adjustments. The uplift and tilting of fault blocks create new river systems which drain along the 'saddle' between adjacent fault blocks, bypassing the basin centre. Most of these bypass rivers finally gain access to the rift axis through transfer zones where the throw on the border fault reduces to zero (see e.g. the Kerio river of northern Kenya described in Frostick and Reid 1987).

In some rifts, topographic barriers pond the drainage, forming lakes. Examples are Lakes Baikal, Tanganyika and Malawi. These lakes vary in salinity from hypersaline to fresh water depending on the surrounding geology and volcanicity. In other rifts there are no lakes and axial rivers drain the length of the valley, for example in the Rhine and Benue rifts. The marginal fault scarps are cut by alluvial fans that feed water and sediment into the basinal rivers and lakes.

As the rift basin floor subsides the uplifted flanks will be progessively eroded and sediments will accumulate in the valley at a rate that largely depends on climate and hydrology. Sediments that accumulate during rifting are normally called 'synrift' sediments. The lowest areas of the valley fill first, generally with lacustrine and river sediment. In the later stages of development into an incipient ocean, sea water may penetrate into the rift valley and the whole area will become a large marine inlet with an uncertain connection to the open ocean. This can lead to the accumulation of thick salt sequences as the sea water evaporates. As the filling progresses wedge-shaped masses of synrift sediments develop and, if subsidence ceased, the valley would eventually lose its topographic identity. In the rifts we see today, subsidence is ongoing and successive wedges of sediment are superimposed on each other. Over geological time many kilometres of sediments can accumulate in rifts.

Continental rifts offer conditions favourable to the development of a number of economic deposits that are rare in other parts of the continents. Some rifts contain sediments that can produce and trap oil and gas in large quantities given the right burial history (e.g. the oil and gas of the North Sea is in a Jurassic rift). Salts that accumulate from both saline lakes and sea water are exploited in some areas for example the Dead Sea Works is situated in the Dead Sea Rift and supplies much of the world's bromine. In addition, the river sands and gravels that accumulate in these basins can be an important source of building materials if the rift is sufficiently close to a developing centre of population.

The spectacular scenery of rifts is, perhaps, their most striking feature and some rifts have therefore become attractive tourist centres. One good example of this is Death Valley in the western USA where the desert conditions reduce vegetation to a minimum and the striking geomorphology is evident even to the untrained eye.

References

Frostick, L.E. (1997) The East African Rift basins, in R.C. Selley (ed.) *African Basins. Sedimentary Basins of the World*, 3, 187–209, Amsterdam: Elsevier.

Frostick, L.E. and Reid, I. (1987) Tectonic controls of desert sediments in rift basins ancient and modern, in L.E. Frostick and I. Reid (eds) *Desert Sediments: Ancient and Modern*, Geological Society Special Publication 35, 53–68.

Frostick, L.E. and Reid, I. (1989) Is structure the main control on river drainage and sedimentation in rifts? *Journal of African Earth Sciences* 8, 165–182.

Girdler, R.W. (1991) The Afro-Arabian rift system – an overview, *Tectonophysics* 197, 139–153.

Summerfield, M.A. (1991) *Global Geomorphology*, 424–425, Harlow: Longman.

Further reading

Frostick, L.E., Renaut, R.W., Reid, I. and Tiercelin, J.J (eds) (1987) *Sedimentation in the African Rift*, Geological Society Special Publication 25, Oxford: Blackwell Scientific.

Hovius, N. and Leeder, M.R. (1998) Clastic sediment supply to basins, *Basin Research* 10, 1–5.

Miall, A.D. (1996) *The Geology of Fluvial Deposits: Sedimentary Facies, Basin Analysis and Petroleum Geology*, Berlin: Springer Verlag.

Selley, R.C. (1997) *African Basins. Sedimentary Basins of the World*, 3, Amsterdam: Elsevier.

LYNNE FROSTICK

RILL

At the start of a rainfall event, rainwater which has fallen upon a hillslope begins to 'pond', i.e. OVERLAND FLOW moves rather slowly under the influence of gravity into small closed depressions in the soil's irregular surface (its 'microtopography'). This 'detention storage' gradually fills, although some of the stored water is constantly lost to infiltration into the soil. Meanwhile, if the rain is of moderate or high intensity then each raindrop which impacts upon an unprotected area of soil will possess sufficient kinetic energy to detach soil particles (see RAINDROP IMPACT, SPLASH AND WASH), which are thus redistributed over the soil's surface. Soil in the ponded areas is however largely protected from raindrop impacts. As a result, rainsplash redistribution usually decreases over time within a storm as the area and depth of surface water increases. There is a net downslope movement of splashed soil but this is generally small.

If the rain continues, then provided precipitation rate exceeds infiltration rate the deepening ponds on the soil's surface will eventually overtop their depressions. Overland flow which is released from overtopped ponds is likely to flow downhill more quickly and in greater quantities (i.e. possess greater kinetic energy) than the more diffuse and shallow flow into depressions: it may therefore be sufficiently competent (see SEDIMENT RATING CURVE) to transport soil particles which are splashed into it. Such soil particles can be carried some distance, being deposited only when flow velocity decreases (due to, for example, a reduction in gradient or the presence of vegetation).

Flow with still greater kinetic energy will generate a shear stress which is sufficient to detach soil particles from the body of the soil. These particles will then be transported along with splashed-in sediment. At locations where such detachment occurs, the soil's surface is lowered slightly. Such lowered areas form preferential paths for subsequent flow, and will thus be eroded further. Rather quickly, this positive feedback (see SYSTEMS IN GEOMORPHOLOGY) results in small, well-defined linear concentrations of flow (Favis-Mortlock 1998), known as 'microrills' or 'traces', with a width and depth of a few millimetres.

Many microrills will eventually become ineffective due to deposition within the microrill itself. But a fortunately located subset may grow further to become rills, with a maximum width and depth of a few tens of centimetres. This process of competition between individual channels leads to the self-organized formation (see COMPLEXITY IN GEOMORPHOLOGY) of networks of microrills and rills. Rill networks tend to be dendritic (see DRAINAGE PATTERN) in form on natural soil surfaces, but are constrained by the direction of tillage on agricultural soils. Such networks form hydraulically efficient pathways for the removal of water from hillslopes. However, sediment which is being transported within the rill network may be redeposited after a short distance if the flow loses its competence (such sediment possibly being detached again later in the same rainfall event, if flow conditions change; or during a subsequent rainfall event). Or the sediment may be carried some distance, perhaps even off the field and into a GULLY and/or a permanently flowing channel (see FIRST-ORDER STREAM). Once the rain stops, however, flow in the rill network will gradually cease: all sediment which is being transported at that time will then be redeposited within the network itself.

Rill networks on agricultural land are regularly erased by tillage, and regularly reinitiated (see SHEET EROSION, SHEET FLOW, SHEET WASH). On natural landscapes, however, rill networks persist and may in time cause such serious dissection of the hillslope as to lead to the formation of BADLANDS.

An eroding hillslope, then, normally consists of a flow-dominated channel network in which rill erosion occurs, separated by interrill areas where the dominant processes are rainsplash and diffuse flow. Soil loss from these areas is known as interrill erosion. But it is rill erosion which is the more effective agent for detachment and removal of soil, and so in many parts of the world rill erosion is the dominant subprocess of SOIL EROSION by water on hillslopes (De Ploey 1983). Boundaries between rill and interrill areas of the hillslope are frequently ill-defined and are constantly shifting. Note that subsurface flow may, in some circumstances, rival hillslope topography in importance in determining where channel erosion will begin and develop, e.g. at the base of slopes, and in areas of very deep soils such as tropical saprolites (see GROUND WATER).

Mean flow velocities within individual rills are usually between one and ten centimetres per second. Interestingly, there is evidence that for actively eroding rills, flow velocity is not dependent on rill gradient: this may be due to some compensatory increase in within-rill roughness on steeper slopes (Nearing *et al.* 1997). Velocity-depth profiles and planform patterns of velocity in rills are qualitatively similar to those of larger channels. Velocities may vary noticeably along the rill, however, with increased flow rates and scouring at 'headcuts' (Slattery and Bryan 1992), i.e. breaks of along-channel slope (such as that which is often found at the upstream extremity of each rill). Headcuts tend to move slowly upstream as their headward facets are eroded.

References

De Ploey, J. (1983) Runoff and rill generation on sandy and loamy topsoils, *Zeitschrift für Geomorphologie N.F. Supplementband* 46, 15–23.

Favis-Mortlock, D.T. (1998) A self-organising dynamic systems approach to the simulation of rill initiation and development on hillslopes, *Computers and Geosciences* 24(4), 353–372.

Nearing, M.A., Norton, L.D., Bulgakov, D.A., Larionov, G.A., West, L.T. and Dontsova, K. (1997) Hydraulics and erosion in eroding rills, *Water Resources Research* 33(4), 865–876.

Slattery, M.C. and Bryan, R.B. (1992) Hydraulic conditions for rill incision under simulated rainfall: a laboratory experiment, *Earth Surface Processes and Landforms* 17, 127–146.

Further reading

Abrahams, A.D., Li, G. and Parsons, A.J. (1996) Rill hydraulics on a semiarid hillslope, southern Arizona *Earth Surface Processes and Landforms* 21(1), 35–47.

Brunton, D.A. and Bryan, R.B. (2000) Rill network development and sediment budgets, *Earth Surface Processes and Landforms* 25(7), 783–800.

Bryan, R.B. (1987) Processes and significance of rill development, *Catena Supplement* 8, 1–15.

Merritt, E. (1984) The identification of four stages during microrill development, *Earth Surface Processes and Landforms* 9, 493–496.

Rauws, G. and Govers, G. (1988) Hydraulic and soil mechanical aspects of rill generation on agricultural soils, *Journal of Soil Science* 39, 111–124.

SEE ALSO: erodibility; erosivity; runoff generation; sheet erosion, sheet flow, sheet wash; soil erosion; universal soil loss equation

DAVID FAVIS-MORTLOCK

RIND, WEATHERING

Weathering rinds are zones of chemical alteration on the outer portions of rocks. In some, but not all cases, a distinct colour difference highlights this zone of intense CHEMICAL WEATHERING. Weathering rinds are important in geomorphology for their role in weathering processes, their role in the development weathering forms such as CASE HARDENING, and in their use in dating landforms. Obsidian-hydration rinds are a related phenomenon.

A weathering rind is not just a zone of chemical alteration at the outer edge of a clast; weathering rinds represent the redistribution of elements. Some rinds are dominated by an enrichment in iron, while others are depleted in such mobile cations as calcium and sodium. A variety of processes develop weathering rinds. Dissolution, for example, leaves void space in the rock and does not necessarily change the colour. Oxidation of iron, in contrast, leaves a band of discolouration. The appearance of the zone of discolouration varies by location and rock type. For instance, rinds can appear white on the upper slopes of Mauna Kea and appear orange on the lower slopes, all in a basalt lithology (Plate 96). Andesite in Japan can appear brown to pale grey (Matsukura *et al.* 1994), and sandstone

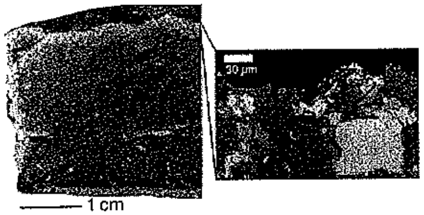


Plate 96 Weathering rind developed on a glacially polished basalt, Mauna Kea, Hawaii. This rind developed over a 16,000-year period. The left photograph shows an optical rind visible in a hand specimen. The right image shows an electron microscope (backscatter) image of a small section of the rind, illustrating three aspects of rind development. First, dissolution of minerals dominates rind formation, as exemplified by the pores (black areas). Second, the bright spots in the image are reprecipitated iron hydroxides, responsible for reddening. Third, rinds may not necessarily thicken over time. Often, they undergo erosion as pieces of weathered minerals progressively detach, that is if rinds are not protected by rock coatings

rinds in New Zealand can appear whitish (Knuepfer 1988).

Weathering rinds form on all three rock types: igneous (e.g. andesites, basalts, granitic), sedimentary (e.g. sandstones) and metamorphic (e.g. schists). Weathering rinds occur in a wide range of locations and in temperate, tropical, arctic and arid environments, for example, Hawaii (Jackson and Keller 1970), the coterminous United States (Colman and Pierce 1986), New Zealand (Chinn 1981), Japan (Matsukura *et al.* 1994) and northern Europe (Dixon *et al.* 2002). Weathering rinds are found in clasts at the surface and within the soil profile (Chinn 1981; Knuepfer 1988).

Weathering rinds are often used in geomorphology to estimate ages of landforms and landscape surfaces (Chinn 1981). This approach assumes that rinds begin to form soon after emplacement of the host rock, and that rinds grow thicker with time (Knuepfer 1988). Weathering rinds thus serve as a relative age indicator where thicker rinds occur on older landforms, and as a calibrated age indicator if accurate forms of age calibration are available in the study area. Prior to the use of cosmogenic nuclides (see COSMOGENIC DATING), use of weathering rinds was prevalent in Quaternary research where moraines, outwash sheets and other landforms correlated climatic changes (Colman and Pierce 1986). The thickness of the discoloured zone of a number of clasts in a deposit is measured normal to the surface, usually with a caliper. Statistical methods differentiate groups of thicknesses among different deposits or surfaces.

Because weathering rinds are so often felt to be synonymous with discolouration, we stress that the study of weathering rinds should not be limited to the measurement of colour changes in hand samples for several reasons. First, a weathering rind can occur without any noticeable colour change. Second, colour change provides only one indication of weathering; microscope studies reveal that the zone of chemical weathering continues into the rock well underneath the zone of colour change. Third, although weathering rinds are not ROCK COATINGs, a single clast may exhibit both a weathering rind and a rock coating (Matsukura *et al.* 1994), a distinction not always recognized in the field. Fourth, where weathering rinds are not protected by rock coatings, weathered mineral fragments readily spall off.

Research into weathering rinds is expanding into exciting new dimensions. Physical and chemical characteristics of weathering rinds are being used to help discern geochemical weathering processes in a given region or area (Dixon *et al.* 2002). The use of cosmogenic nuclides as a dating method has made weathering rind analysis more important than ever. A key uncertainty in cosmogenic dating surrounds the prior exposure history of a possible sample. With each cosmogenic measurement costing about US$2,000 in sample processing and analysis, weathering-rind measurements provide an inexpensive field check on the possibility that a particular sample might have a complex geomorphic history. In addition, *in situ* measurements of weathered minerals in rinds are providing new insight into quantitative rates of weathering; this method is being used, for example, to establish long-term rates of glass dissolution with the goal of understanding GEOMORPHOLOGICAL HAZARDs associated with nuclear waste storage (Gordon and Brady 2002).

References

Chinn, T. (1981) Use of rock weathering-rind thickness for Holocene absolute age-dating in New Zealand, *Arctic and Alpine Research* 13, 33–45.

Colman, S.M. and Pierce, K.L. (1986) Glacial sequence near McCall, Idaho: weathering rinds, soil development, morphology, and other relative-age criteria, *Quaternary Research* 25, 25–42.

Dixon, J.C., Thorn, C.E., Darmody, R.G. and Campbell, S.W. (2002) Weathering rinds and rock coatings from an Arctic alpine environment, northern Scandinavia, *Geological Society of America Bulletin* 114, 226–238.

Gordon, S.J. and Brady, P.V. (2002) In situ determination of long-term basaltic glass dissolution in the unsaturated zone, *Chemical Geology* 90, 115–124.

Jackson, T.A. and Keller, W.D. (1970) A comparative study of the role of lichens and 'inorganic' processes in the chemical weathering of recent Hawaiian lava flows, *American Journal of Science* 269, 446–466.

Knuepfer, R.L.K. (1988) Estimating ages of late Quaternary stream terraces from analysis of weathering rinds and soils, *Geological Society of America Bulletin* 100, 1,224–1,236.

Matsukura, Y., Kimata, M. and Yokoyama, S. (1994) Formation of weathering rinds on andesite blocks under the influence of volcanic gases around the active crater Aso Volcano, Japan, in D.A. Robinson and R.B.G. Williams (eds) *Rock Weathering and Landform Evolution*, 89–98, Chichester: Wiley.

SEE ALSO: case hardening; chemical weathering; rock coating

STEVEN J. GORDON AND RONALD I. DORN

RING COMPLEX OR STRUCTURE

A petrologically variable but structurally distinctive group of hypabyssal or subvolcanic igneous intrusions that include ring dykes, partial ring dykes and cone sheets. Outcrop patterns are arcuate, annular, polygonal and elliptical with varying diameters ranging from less than 1 to 30 km or greater. The majority of ring complexes represent the eroded roots of volcanoes and their calderas.

(Bowden 1985: 17)

Ring dykes are thick, approximately vertical igneous bodies that form concentric circles around a central intrusion. They are associated with a process called cauldron subsidence. Cone sheets tend to be thinner and have a general form as a set of inverted cones. They result from stresses set up in the Earth's crust as the magma body with which they are associated forced its way upwards. Other circular structures are associated with impact events.

Reference

Bowden, P. (1985) The geochemistry and mineralization of alkaline ring complexes in Africa (a review), *Journal of African Earth Sciences* 3, 17–39.

A.S. GOUDIE

RIP CURRENT

Many of the world's BEACHes are characterized by the presence of strong, concentrated seaward flows called rip currents. The term was introduced by Shepard (1936) to distinguish rips from the misnomers 'rip tide' and 'undertow', which are unfortunately often still used to describe rips today. Rips are an integral component of nearshore cell circulation and ideally consist of two converging longshore feeder currents which meet and turn seawards into a narrow, fast-flowing rip-neck that extends through the surf zone, decelerating and expanding into a rip-head past the line of breaking WAVEs. The circulation cell is completed by net onshore flow due to wave mass transport between adjacent rip systems (Figure 133a). Rip flows are often contained within distinct topographic channels between bars (see BAR, COASTAL) and are a major mechanism for the seaward transport of water, sediments and pollutants (Figure 133b). Rips are also a major hazard to swimmers and it is of concern that many aspects of rip occurrence, generation and behaviour remain poorly understood.

Rip currents are generally absent on pure dissipative and reflective beaches, but are a key component of sandy intermediate beach states in microtidal environments. Short (1985) identified three types: (1) accretion rips occur during decreasing or stable wave energy conditions and are often topographically arrested in position with mean velocities typically on the order of $0.5-1\,\mathrm{m\,s^{-1}}$; (2) erosion rips are hydrodynamically controlled and occur under rising wave energy conditions. They are transient in location, having mean flows in excess of $1\,\mathrm{m\,s^{-1}}$; and (3) mega-rips, which occur in embayments under extremely high waves, and can extend more than 1 km offshore with mean velocities greater than 2 $\mathrm{m\,s^{-1}}$. All are associated with localized erosion of the shoreline and often create rhythmic embayments termed mega-cusps. Relatively permanent rips located adjacent to headlands, reefs and coastal structures, such as GROYNEs, are referred to as topographically controlled rips.

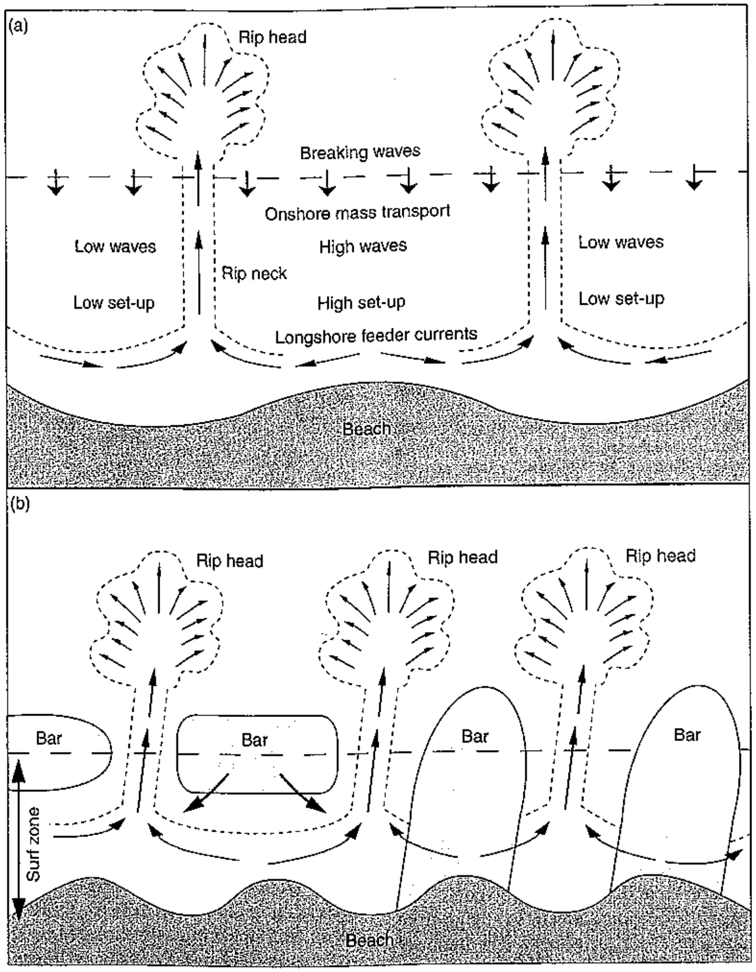


Figure 133 Idealized patterns of rip current flow and components in relation to: (a) nearshore cell circulation and wave set-up gradients; and (b) coastal bar topography

The primary limitation to our understanding of rips has been the difficulty obtaining quantitative field measurements from an energetic environment. Early attempts at describing rips (e.g. McKenzie 1958) were largely qualitative, but correctly identified that rips often display a periodic longshore spacing, increase in intensity and decrease in number as wave height increases, and flow fastest at low tide. Subsequent theoretical, laboratory and field studies have attempted to explain these characteristics with varying degrees of success, although it is generally accepted that rips exist as a response to an excess of water, termed wave set-up, built up on shore by breaking waves. The flow is forced by longshore variations in wave height, which produce gradients in the set-up that drive water alongshore from regions of high to low waves (Bowen 1969; see Figure 133a).

Existing models for the generation of rip cell circulation have thus incorporated various mechanisms to account for the existence of these longshore gradients and can be grouped into three main categories: (1) the wave–boundary interaction model involves wave modification by non-uniform topography and/or coastal structures. For example, wave refraction can produce regions of high and low waves, such that rips can occur in the lee of offshore submarine canyons (Shepard and Inman 1950), but more commonly adjacent to headlands and groynes; (2) wave–wave interaction models have shown theoretically and in laboratory experiments (Bowen and Inman 1969) that incident waves can generate synchronous edge waves that produce alternating patterns of high and low wave heights along the shoreline. Rips occur at every other antinode with a spacing equal to the edge wave length; and (3) instability models suggest that longshore uniformity in set-up is unstable to any small disturbance caused by hydrodynamic or topographic factors and rip spacing is predicted to equal four times the surf zone width. It should be emphasized that validation of these models has primarily been restricted to laboratory experiments and has not been adequately verified in the field. Short and Brander (1999) used a global field dataset to show that rip spacing is related to regional wave energy environments. Patterns of rip spacing (L_r) were consistent within west coast swell ($L_r \cong 500$ m), east coast swell ($L_r \cong 200$ m), and fetch-limited wind wave environments ($L_r \cong 50-100$ m).

The wave–boundary model is best supported by Sonu (1972) who found that on a beach consisting of alternating sandbars and topographic channels with uniform longshore wave height, constant and extensive wave energy dissipation across the bars and local and intense wave breaking over the channels created a set-up gradient towards the channels, which controlled rip flow. Set-up gradients generated in this manner support field data confirmation (e.g. Brander 1999) that rip flows are tidally modulated, since stronger flows at low tide would be expected with increased wave dissipation associated with shallower water depths over the bars. Field studies have also shown that rip velocities increase steadily from the feeders, attaining maximums in the middle of the rip-neck, are greater near the water surface and experience short duration and strong velocity pulses every few minutes, the forcing of which is likely related to infragravity motions such as shear waves or wave groups.

References

Bowen, A.J. (1969) Rip currents. 1. Theoretical investigations, *Journal of Geophysical Research* 74, 5,467–5,478.

Bowen, A.J. and Inman, D.L (1969) Rip currents. 2. Laboratory and field observations, *Journal of Geophysical Research* 74, 5,479–5,490.

Brander, R.W. (1999) Field observations on the morphodynamic evolution of a low-energy rip current system, *Marine Geology* 157, 199–217.

McKenzie, P. (1958) Rip-current systems, *Journal of Geology* 66, 103–111.

Shepard, F.P. (1936) Undertow, rip tide, or rip current, *Science* 84, 181–182.

Shepard, F.P. and Inman, D.L. (1950) Nearshore water circulation related to bottom topography and wave refraction, *Transactions of the American Geophysical Union* 31, 196–212.

Short, A.D. (1985) Rip current type, spacing and persistence, Narrabeen Beach, Australia, *Marine Geology* 65, 47–61.

Short, A.D. and Brander, R.W. (1999) Regional variations in rip density, *Journal of Coastal Research* 15(3), 813–822.

Sonu, C.J. (1972) Field observation of nearshore circulation and meandering currents, *Journal of Geophysical Research* 77, 3,232–3,247.

SEE ALSO: bar, coastal; beach; beach sediment transport; groyne; wave

ROBERT W. BRANDER

RIPARIAN GEOMORPHOLOGY

Riparian geomorphology is concerned with the dynamics, form and sedimentary structure of riparian zones. Riparian zones have been variously described as 'three-dimensional zones of direct interaction between terrestrial and aquatic ecosystems' (Gregory *et al.* 1991: 540); zones that extend 'from recently colonized fluvial landforms exposed at low flow to the limits of the area wherein biota are adapted to, or characteristic community structures are influenced by, flooding' (Dykaar and Wigington 2000: 88); and the 'part of the biosphere supported by, and including, recent fluvial landforms. . .inundated or saturated by the BANKFULL DISCHARGE' (Hupp and Osterkamp 1996: 280). From these and other definitions, it is apparent that the riparian geomorphology of a river reach is dependent upon: present and past flow magnitude and frequency (see MAGNITUDE–FREQUENCY CONCEPT); the amount and calibre of sediment transported by the river; and the slope and degree of confinement of the reach. Whilst the past and present flow and sediment transport regime govern the materials delivered to the reach for landform building, the local slope and confinement of the reach govern the river's energy and its ability to construct and erode landforms.

Nanson and Croke (1992) explored these controls to develop a genetic classification of FLOODPLAIN types that is explicitly linked to the river types that construct the floodplains. They defined three broad groups of floodplain (high-energy non-cohesive, medium-energy non-cohesive, low-energy cohesive) based upon the river's ability to do work and expressed by its specific STREAM POWER at bankfull discharge, and the erosional resistance of the floodplain materials (non-cohesive implies gravel or sand; cohesive implies silt and clay). They subdivided the three groups into thirteen different floodplain classes. These were discriminated by details of the sediment from which they are constructed, the river planform or pattern, its characteristic erosional and depositional processes, and thus the typical landforms present on the floodplain and within the river margins. Importantly, this classification links process and form in a dynamic way, illustrating that riparian zones may possess an enormous variety of landforms and that the nature and dynamism of the landforms varies between floodplain types. Thus, if there is a change in the controlling processes, riparian geomorphology also changes. For example, changes in climate, flow regulation and flood defence engineering affect river flow and sediment transport regimes and the ERODIBILITY of channel margin materials, and so can have far-reaching impacts on riparian zone character (e.g. Steiger and Gurnell 2002).

In most analyses of riparian zone form and process, vegetation has been seen to play a largely passive role, responding to present and past environmental conditions created by fluvial processes (e.g. Hupp 1988). Thus, floodplain vegetation patterns have been interpreted to depend on the type and age of the mosaic of riparian landforms. Migrating, MEANDERING rivers provide a simple illustration. As the river erodes the outer banks of meander bends, POINT BARs develop on the inner banks. Vegetation colonizes point bar surfaces and plant species are gradually replaced as sediment, moisture, light and disturbance on the bars change during their aggradation and incorporation into the floodplain.

Recently, more emphasis has been placed on the active role of vegetation in influencing riparian zone geomorphology. For example, Gurnell and Petts (2002) consider both biotic and abiotic ways in which vegetation can influence the form, sedimentary structure and dynamics of riparian zones. Abiotic influences include the impact of root systems on the erodibility of sediments and the flow resistance of the vegetation canopy. Roots can cause significant reinforcement of riparian sediments, making them more resistant to river erosion. When the riparian zone is flooded, the ROUGHNESS of the vegetation canopy can reduce flow velocities across the vegetated surface, reducing rates of erosion and increasing rates of sedimentation. These abiotic processes can significantly affect patterns of erosion and aggradation, and thus the form and sedimentary structure of riparian zones. The geomorphological significance of these abiotic influences depends on the species, age and density of the vegetation cover, which is related to several biotic processes. The degree to which riparian plants reproduce from seeds or by vegetative reproduction is important because, in general, riparian vegetation growth is more rapid when plants propagate vegetatively. The timing of seed or vegetative propagule release can greatly influence the likelihood of successful vegetation establishment, because many riparian plant propagules are transported and deposited by the river. For example, the timing of propagule release in relation to the climate and river flow regimes can influence whether suitable colonization sites are exposed or inundated by the river, and whether their moisture and temperature characteristics are appropriate to support the successful germination and growth of young plants.

Riparian tree species can be particularly important riparian zone engineers. Poplar and willow species can grow very rapidly, propagating through both seeds and vegetative reproduction. Rivers may erode, transport and deposit whole trees as well as fragments (branches, twigs, root boles) and seeds. Entire trees may survive rafting by floods, deposition and burial within river margins and on bars, and they can sprout to form patches of new sizeable shrubs within a year. The importance of these processes for riparian geomorphology varies with tree species and environmental conditions but also with riparian tree management. The pruning and felling of riparian trees to prevent LARGE WOODY DEBRIS entering rivers is often carried out to

maintain the FLOOD conveyance of the river channel. Its impact on riparian geomorphology and ecology is far reaching, leaving little impression of the diverse geomorphological and ecological character and high dynamism of unimpacted riparian zones (Gurnell *et al.* 1995, 2002).

References

Dykaar, B.B. and Wigington, P.J. (2000) Floodplain formation and cottonwood colonization patterns of the Willamette River, Oregon, USA, *Environmental Management* 25, 87–104.

Gregory, S.V., Swanson, F.J., McKee, W.A. and Cummins, K.W. (1991) An ecosystem perspective of riparian zones *BioScience* 41, 540–551.

Gurnell, A.M., Gregory K.J. and Petts G.E. (1995) The role of coarse woody debris in forest aquatic habitats: implications for management, *Aquatic Conservation* 5, 143–166.

Gurnell, A.M. and Petts, G.E. (2002) Island-dominated landscapes of large floodplain rivers, a European perspective, *Freshwater Biology* 47, 581–600.

Gurnell, A.M., Piégay, H., Swanson, F.J. and Gregory, S.V. (2002) Large wood and fluvial processes, *Freshwater Biology* 47, 601–619.

Hupp, C.R. (1988) Plant ecological aspects of flood geomorphology, in V.R. Baker, R.C. Kochel and P.C. Patten (eds) *Flood Geomorphology*, 335–356, New York: Wiley.

Hupp, C.R. and Osterkamp, W.R. (1996) Riparian vegetation and fluvial geomorphic processes, *Geomorphology* 14, 277–295.

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

Steiger, J. and Gurnell, A.M. (2002) Spatial hydrogeomorphological influences on sediment and nutrient deposition in riparian zones: observations from the Garonne River, France, *Geomorphology* 49(1), 1–23.

Further reading

Gurnell, A.M., Hupp, C.R. and Gregory, S.V. (eds) (2000) Linking hydrology and ecology, *Hydrological Processes*, Special Issue 14, 2,813–3,179.

Stanford, J.A. and Gonser, T. (eds) (1998) Rivers in the landscape: riparian and groundwater ecology, *Freshwater Biology*, Special Issue 40, 401–585.

Tockner, K., Ward, J.V., Kollmann, J. and Edwards, P.J. (eds) (2002) Riverine Landscapes, *Freshwater Biology*, Special Issue 47, 497–907.

ANGELA GURNELL

RIPPLE

Ripple is a general term applied to a range of normally unrelated, very small bedforms that occur in trains and record sediment mobilization and transport in various aqueous and aeolian environments (see BEDFORM; BEDLOAD; ROUGHNESS). The main kinds are current and rhomboid ripples, oscillation or 'symmetrical' ripples (see WAVE), ballistic or impact ripples (see AEOLIAN PROCESSES; SALTATION), adhesion ripples and warts (see AEOLIAN PROCESSES; SALTATION), and rain-impact ripples (see RAINDROP IMPACT, SPLASH AND WASH). With the exception of adhesion warts, and some complex oscillation types, ripples are characterized by crests that lie transversely to flow.

Trains of current ripples, restricted to the coarser silts and the finer sands, are typical of rivers, but also appear in tidal environments (estuaries, barred beaches) where flows can be unidirectional for several hours at a time. As equilibrium bedforms, current ripples have linguoid crests in plan, heights of up to about 0.02 m, wavelengths of 0.1–0.2 m, and strongly asymmetrical profiles, the short leeward face lying at the angle of repose (see REPOSE, ANGLE OF). When generated from a smooth bed, however, current ripples evolve toward a linguoid shape through a range of long-crested forms, the crests of which increasingly lose straightness. Internally, current ripples are cross-laminated, commonly in climbing sets, a testimony to high rates of sediment deposition on a scale of minutes or hours. Ripple dimensions are independent of flow depth but increase weakly with grain size. Diamond-shaped rhomboid ripples are developed where ripple-generating flows are sufficiently shallow as to be supercritical.

Other flows being absent, wind waves generate within the affected water-body symmetrical, oscillatory currents which are superimposed on a much weaker drift in the direction of wave-propagation. When sufficiently powerful, their combined effect on sand beds is to create trains of ripples with long, regular crests and steep, almost symmetrical, trochoidal profiles which, as revealed by internal cross-laminae, migrate very slowly in the direction of wave-propagation. Ripple scale depends in a complex manner on the properties of the waves and the sediment, the wavelength and height increasing markedly with grain size. Wavelengths are of the order of 0.01 m in silt, 0.1 m in fine sand and 1 m in coarse sands and fine gravels. Broadly, wavelength is about 500 times the median grain diameter. Wave ripples are most familiar from estuaries and beaches but, after storms, appear on continental shelves to water depths of 100–200 m. Complex forms of ripple occur where barriers reflect waves and where, especially on beaches and in estuaries, unrelated unidirectional and wave currents operate either simultaneously or sequentially. Wave ripples are valuable indicators of shallow water and of shoreline location and orientation.

The saltation of wind-driven grains over a dry bed is generally accompanied by the development of trains of ballistic ripples, resulting from an unstable interaction between the surface and the flow of sediment. These ripples are rather flat, asymmetrical structures which vary in form and scale with increasing grain size and the average length of the jumps made by the particles. Typically, ripples in the finer sands have crests that are long and regular in plan and wavelengths of about 0.05 m. Those in sediments of very coarse sand or granule grade take wavelengths of the order of 1 m and generally have short, irregular crests, along which the coarser particles conspicuously lie. Ballistic ripples are cross-laminated internally, but the structure is difficult to see in the well-sorted sands of which the smaller examples are formed. The ripples have long been reported from deserts and sandy coasts, wherever the wind is free to mobilize sufficiently coarse grains.

The capture of saltating particles by a damp or wet surface, such as a coastal sand beach, river bar or sabkha, gives rise to upwind-facing, centimetre-scale adhesion ripples (uniform wind-direction) or adhesion warts (wind-direction variable). These common and widespread structures have no particular climatic significance but are valuable proofs of surface exposure and aeolian activity. Advancing in the opposite direction to the wind, adhesion ripples create a steep internal bedding that dips downwind.

Rain-impact ripples are centimetre-scale, upwind-facing, transverse ridges shaped when heavy rain driven by a strong wind descends at a fine angle onto an exposed, water-saturated sand bed, such as a beach, tidal sand shoal or river bar. The ridges advance very slowly in the direction of the wind under the repeated impact of the drops. If rain-impact ripples have a fossil record, which is uncertain, they would afford a further proof of atmospheric exposure.

Further reading

Allen, J.R.L. (1979) A model for the interpretation of wave ripple-marks using their wavelength, textural composition and shape, *Journal of the Geological Society, London* 136, 673–682.

——(1982) *Sedimentary Structures*, Amsterdam: Elsevier.

Anderson, R.S. (1987) A theoretical model for aeolian impact ripples, *Sedimentology* 34, 943–956.

Baas, J.H. (1999) An empirical model for the development of and equilibrium morphology of current ripples in very fine sand, *Sedimentology* 46, 123–138.

Bagnold, R.H. (1946) Motion of waves in shallow water. Interaction between waves and sand bottom, *Proceedings of the Royal Society, London* A187, 1–16.

Clifton, H.E. (1977) Rain-impact ripples, *Journal of Sedimentary Petrology* 47, 678–679.

Doucette, J.S. (2002) Geometry and grains-size sorting of ripples on low-energy sandy beaches; field observations and model predictions, *Sedimentology* 49, 483–503.

Fryberger, S.G., Hesp, P. and Hastings, K. (1992) Aeolian granule ripple deposits, Namibia, *Sedimentology* 39, 319–331.

Kahle, C.F. and Livchak, C.J. (1996) Nature and significance of rhomboid ripples in a Silurian sabkha sequence, north-central Ohio, *Journal of Sedimentary Research* 66, 861–867.

Kocurek, G. and Fielder, G. (1982) Adhesion structures, *Journal of Sedimentary Petrology* 52, 1,229–1,241.

J.R.L. ALLEN

RIVER CAPTURE

River capture, sometimes called stream capture or stream piracy, refers to the occurrence of the seizure of the waters of a stream or drainage system by a neighbouring one. It is based on the difference in local BASE LEVEL heights, with the captured stream having a higher base level and for that reason with a low erosion potential. The predatory stream, with a lower base level, is capable of diverting in its favour the waters of the less active stream, and in this way enlarging its drainage net and catchment area. Integration of both drainage systems leads to a higher order network. It does not only occur because of a steeper gradient but also because the pirate stream is cutting its valley in softer rock.

Capture constitutes a common event in the erosional evolution of a drainage net of a region and is a traditional concept in geomorphology that can be found in classical authors. Gilbert (1877) described the process in relation to the role of unconsolidated materials in mine dumps, calling it abstraction, a term often applied to the simplest type of capture, which results from competition between adjacent consequent gullies and ravines. He was also aware that a stream flowing down the steeper slope of an asymmetrical ridge erodes its valley more rapidly than the one flowing down a more gentle slope, and as a result the divide migrates away from the more actively eroding stream. This principle has been refered to as the *law of unequal slopes* (Thornbury 1969). The same concept was integrated by Davis (1899) in his model of relief evolution by the geographical cycle, capture taking place in the young or early mature stages of development. Another classical author, Horton (1945), in his slope runoff model also takes into consideration the capture process and uses it in order to explain the development of a hierarchical drainage net, that is, the process by which the drainage lines become integrated into a few dominant stream courses. Unequal rainfall on two sides of a divide may contribute to divide migration, especially where winds are prevailing from one direction, as in the trade wind belts (Thornbury 1969).

At the point at which the capture takes place, the captured stream bends sharply, forming a right angle turn into the pirate stream, which is called the elbow of capture. The valley stretch in which the captured stream continues to flow after losing the upper part of its catchment becomes a beheaded valley. This valley is then too large for the stream that continues to flow in it and thus becomes an UNDERFIT STREAM, that is a stream too small to be hydrologically related to the valley in which it now flows. On the other hand, the captured part of the stream now has a lower local base level, which increases its erosion potential and makes it able to incise into its former alluvial valley floor, producing a terrace in its former floodplain.

The capture process mainly occurs in two different ways: by headward erosion and by lateral erosion. *Headward erosion* is the probable cause of most easily recognizable stream captures. It takes place when the tributaries of the high energy stream are working back towards its head, and eventually reach the neighbouring valley head and cut through the divide. Capture by *lateral erosion* occurs when two streams flow parallel at no great distance from each other. Progressive erosion produces a lateral shifting of the stream which can finally produce a planation of the water divide. If this continues at the cost of one of the neighbouring streams, it ends in the lateral capture of its waters. Capture by subterranean waters can also occur in soluble rocks when water from a stream at a higher level percolates and meets an underground stream flowing at a lower level.

Examples of river captures have been described in many regions of the world, both at large and at small scales. In the large scale, one of the classical examples is that of a tributary of the Indus captured by the Ganges which implied the transfer of the drainage of a large area of the Himalayas from Pakistan to India. In Yunnan Province, China, rivers flowing towards the Red river were captured by the middle Yangtze tributaries. In Queensland, eastern Australia, the Fitzroy River has reached an old divide at the Connors Range and captured several of the rivers flowing in this area. In New Zealand the capture of the Silver Stream by the Karori near Wellington is well known. In Europe waters were diverted from the Danube towards the Rhine by a small head-ward tributary. In North America, in the Appalachian region of the eastern United States, there are many captures which are controlled by differences in rock hardness.

Amongst the implications of river captures is their role and significance for the evolution of relief and the history of drainage patterns, providing an interesting geomorphic challenge. In that sense, a highly integrated stream system with a large main stream is usually an indication of a long period of development (Ahnert 1998). Another important issue mentioned by Schumm (1977) is the role of captures in the discovery of new placer deposits, which are alluvial deposits containing valuable minerals, because the regional distribution of placers can be strongly influenced by stream capture. The result of this process from an economic point of view is that the source of the valuable minerals may be abruptly isolated from the downstream depositional area.

References

Ahnert, F. (1998) *Introduction to Geomorphology*, London: Arnold.

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

Gilbert, G.K. (1877) Report on the geology of the Henry Mountains, 141, Washington, *US Geographical and Geological Survey of the Rocky Mountains Region.*

Horton, R.E. (1945) Erosional development of streams and their drainage basins: hydrological application of quantitative morphology, *Geological Society of America Bulletin* 56, 281–370.

Schumm, S.A. (1977) *The River System*, Chichester: Wiley.

Thornbury, W.D. (1969) *Principles of Geomorphology*, New York: Wiley.

SEE ALSO: base level; gully; underfit stream

MARIA SALA

RIVER CONTINUUM

The biological concept of a river continuum describes a regular downstream progression of such physical variables as channel width, diel temperature pulse and stream order, in relation to biotic adjustments (Vannote *et al.* 1980). The concept was originally developed for rivers in regions with deciduous forests. In these regions, headwater streams (orders 1–3) are narrow and shaded by riparian vegetation. The vegetation reduces instream or autotrophic production from algae by shading, and contributes large amounts of coarse organic detritus (>1 mm diameter), such as leaf litter. The ratio of photosynthesis/respiration (P/R) is less than 1. The diversity of soluble organic compounds is high, and the diel temperature pulse is low. Communities of aquatic insects in headwater streams are dominated by insects that shred coarse organic matter (shredders), and insects that filter finer organic matter from transport, or gather such material from sediments (collectors). Fish populations have cool-water species that feed mainly on invertebrates. Biotic diversity is low.

Medium-sized streams (orders 4–6) are sufficiently wide that sunlight reaches a greater portion of the stream channel. Algae and rooted plants in the stream are more plentiful, and the ratio of P/R exceeds 1. The diversity of soluble organic compounds drops sharply relative to headwater streams, and the diel temperature pulse reaches a maximum. Fine particulate organic matter (50 μm–1 mm diameter) becomes more important. Collectors remain important, shredders form a smaller percentage of insect communities, and grazers that shear attached algae from surfaces in the stream increase in abundance. Fish populations now have more warm-water species that feed on invertebrates and other fish. Biotic diversity reaches a maximum.

Large rivers (>6 order) are very broad and open to sunlight, but photosynthesis may be limited by depth and turbidity. Large quantities of fine particulate organic matter from processing of dead leaves and woody debris come from upstream, and the ratio of P/R again drops below 1. The diel temperature pulse is low. Aquatic insects are primarily collectors. Fish are warm-water species that feed on plankton, invertebrates and other fish. Biotic diversity drops off again.

The general pattern described above may differ in mountainous areas where headwater streams flow through alpine meadows, in dry regions where riparian vegetation is restricted, or along deeply incised channels where shading from valley walls limits photosynthesis. However, the river continuum does provide a conceptual model of spatial gradients in physical and biological variables. This conceptual model is one of the first holistic theories of a river as an ecosystem, rather than individual segments. The river continuum emphasizes the connections between the river and its terrestrial setting. The continuum also suggests that aquatic communities can be explained by the mean state of environmental variables and their degree of temporal variability and spatial heterogeneity (Minshall *et al.* 1985). Together with hypotheses of stream succession that predict changes in habitat and species following a disturbance such as a FLOOD, the river continuum concept facilitates predictions of reach-specific patterns in habitat, communities or life-history strategies.

References

Minshall, G.W., Cummins, K.W., Petersen, R.C., Cushing, C.E., Bruns, D.A., Sedell, J.R. and Vannote, R.L. (1985) Developments in stream ecosystem theory, *Canadian Journal of Fisheries and Aquatic Science* 42, 1,045–1,055.

Vannote, R.L., Minshall, G.W., Cummins, K.W., Sedell, J.R. and Cushing, C.E. (1980) The river continuum concept, *Canadian Journal of Fisheries and Aquatic Science* 37, 130–137.

SEE ALSO: fluvial geomorphology; large woody debris; stream ordering

ELLEN E. WOHL

RIVER DELTA

River deltas are coastal accumulations of terrestrial sediments that rivers have brought to the sea. The Greek historian Herodotus (*c.*450 BC) originally applied the term 'delta' to the triangular subaerial deposit surrounding the mouth of the Nile River. In modern usage, however, deltas can be either subaerial or subaqueous accumulations and may have a variety of geometries. Although deep sea fans may also be considered deltas, they are not discussed here. Here, the 'subaqueous delta' is assumed to be confined to deposits on the continental shelf. The prevailing shape of any given delta depends on the rates of sediment supply by the rivers and the patterns and rates of sediment

dispersal by coastal ocean processes and by gravity. In many cases, the subaqueous deltaic deposits are much more extensive than the subaerial deposits and, in some cases such as that of Papua New Guinea's Sepik River which discharges directly into deep water, the subaerial delta may be missing altogether. Historically, deltas have played important socio-economic roles. Subaerial deltas were the sites of early agriculture and formative civilizations and presently support some of the world's largest urban centres (e.g. Shanghai, Bangkok, Cairo). Subaqueous deltas are sinks for terrestrial carbon and are sources of fossil fuel.

Deltas vary immensely in both area and volume. The size of a delta depends at the lowest order on the annual sediment discharge of the river but the most extensive deltas also tend to be developed where wide, low gradient continental shelves provide a platform for prolonged sediment accumulation and morphological progradation. Hence, the largest deltas are found on passive (as opposed to active) continental margins (Wright 1985). Despite this fact, active margins are probably equally or more important than passive margins in supplying river sediment to the sea; Milliman and Syvitski (1992) showed that the numerous small mountainous streams, particularly those of the humid tropics, are collectively the most important source of terrestrial sediment to the sea. However, since these rivers are spatially distributed and since much of this sediment is bypassed to deep water, large deltas typically do not result. Other factors that influence delta area and the relative sizes of subaerial vs subaqueous components include tectonic subsidence and the energy of waves and currents that resuspend or prevent shallow water deposition of sediments. Table 38 lists characteristics of five deltas.

Satellite images of the Changjiang (Yangtze) and Mississippi Deltas (Plates 97 and 98) illustrate the diversity of major deltas. In the case of the Changjiang (Plate 97), the river-supplied sediments have built a large lobate protrusion into the East China Sea that supports and surrounds Shanghai. More exaggerated seaward protrusion of this delta is constrained by strong currents and waves, which disperse newly discharged sediments over the shelf and into Hangzhou Bay. These sediments are accumulating on the shelf at a rate of $5\,\mathrm{cm\,yr^{-1}}$ (DeMaster *et al.* 1985) to form the subaqueous component of the delta. The Mississippi Delta, one of the world's most extensively studied deltas, is composed of sediments from a catchment that covers 60 per cent of the continental United States. Its elongated and narrow digitate shape is rare and is attributable to a combination of fine cohesive sediment, low wave

Table 38 Characteristics of five major river deltas

Property	River delta				
	Amazon	Ganges–Brahmaputra	Changjiang (Yangtze)	Huanghe (Yellow R.)	Mississippi
Drainage Basin Area ($\mathrm{km^2} \times 10^6$)	6.15	1.48	1.94	0.77	3.27
Water discharge ($\mathrm{km^3\,yr^{-1}}$)	6,300	971	900	42	580
Sediment discharge (10^6 tonnes $\mathrm{yr^{-1}}$)	900	1,620	480	1,060	210
Sediment/water ratio ($\mathrm{kg\,m^{-3}}$)	0.14	1.67	0.53	25.25	0.36
RMS wave height (m)	1.6	2.5	1.5	2.0	1.1
Spring tide range (m)	5.8	3.6	3.0	1.4	0.4
Total delta area ($\mathrm{km^2} \times 10^3$)	467	106	67	36	29
Ratio subaerial/ subaqueous area	6.4	2.4	1.7	3.3	5.3

Sources: Coleman and Wright (1975); Milliman and Meade (1983); Wright and Nittrouer (1995)

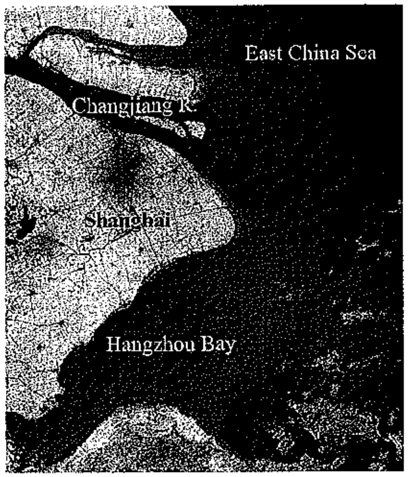


Plate 97 Satellite image of the Changjiang (Yangtze) delta and estuary showing the city of Shanghai, the turbid river effluent and Hangzhou Bay which serves as a sink for much of the sediment

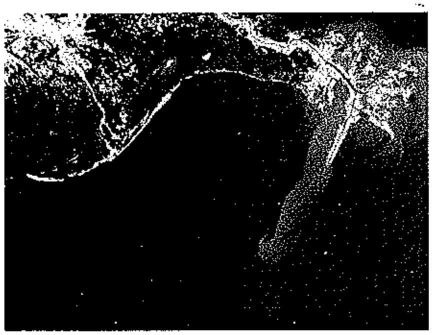


Plate 98 Satellite image of the Mississippi Delta showing the active 'bird's foot' in the right portion of the image and the abandoned La Fourche delta lobe on the left

height and negligible tidal currents, a regime that allows sediments to accumulate near the river mouths without being widely dispersed by oceanographic forces (Wright and Nittrouer 1995). The modern Mississippi 'bird's foot' (Plate 98) now extends across the continental shelf creating a barrier to east–west currents. In recent geologic history, a series of such lobate overextensions have been followed by avulsions: at least sixteen such lobes have been created and abandoned in Holocene time (Kolb and Van Lopik 1966). Abandoned deltaic lobes make up most of Louisiana's coastal plain, which is experiencing a high rate of coastal land loss because of regional subsidence and erosion.

The processes that disperse, transport and deposit the sediment discharged by a river determine the configuration of the resulting delta. This is true not only for the subaqueous component but also for the subaerial delta, which must surmount the subaqueous deposits in order to prograde. Wright and Nittrouer (1995) argued that the fate of sediment seaward of river mouths involves at least four stages: (1) supply via river plumes; (2) initial deposition; (3) resuspension and onward transport by marine forces (e.g. waves and currents); and (4) long-term net accumulation. Different suites of processes dominate each stage. Immediately upon leaving the confines of a river mouth, a river effluent spreads as either a positively buoyant (lighter than seawater because of the salinity difference) or negatively buoyant (because of very high suspended sediment concentrations in the river water) plume while mixing and exchanging momentum with the ambient seawater. This is the first stage of dispersal. Most of the larger rivers that drain to passive continental margins have positively buoyant effluents because they carry low concentrations of suspended sediment and large volumes of low-density fresh water. Examples include the Mississippi, Amazon, Ganges–Brahmaputra and Changjiang among numerous others. The most prominent exception among the large rivers is the Huanghe (Table 38), which often transports suspended sediment in concentrations greater than $25\,kg\,m^{-3}$ creating an effluent bulk density greater than that of seawater (Wright *et al.* 1990). Such negatively buoyant effluents are referred to as hyperpycnal (excessively dense) and tend to move downslope within the near-bed layer under the influence of gravity. Hyperpycnal conditions are somewhat more common at the mouths of smaller rivers that drain mountainous catchments near the coasts of active margins. The Eel River of northern California is a prominent example (Geyer *et al.* 2000).

The second stage of sediment dispersal is represented by initial, but usually temporary, deposition from the expanding and decelerating effluents of stage 1. River-mouth bars of varying geometries are among the morphologic products of this deposition. The deposition is caused by sediment flux convergence produced by effluent deceleration and sediment particle settling. The more rapidly the effluent gives up its momentum through mixing and bed friction and the greater the particle setting velocity, the closer the one would expect initial deposition to be to the river mouth. On the other hand, high waves and strong coastal currents enhance mixing and effluent momentum exchange with the sea but may also act to resuspend sediment or to inhibit initial deposition. Along high-energy coasts, the initial deposition may be delayed until the sediment reaches a deeper, more quiescent environment such as the mid-shelf region (Ogston *et al.* 2000). Energetic oceanographic processes also disperse sediment parallel to the coast or isobaths, generally preventing the formation of digitate or protruding deposits.

The third, resuspension and transport, stage of dispersal may act concurrent with or subsequent to the initial deposition stage (stage 2). When the coastal regime is highly energetic throughout the year or when high energy coincides with high river discharge, deposition is delayed as explained above. However, in many cases (e.g. Huanghe; Wright *et al.* 1990), the maximum input of river sediments to the sea and the maximum agitation of the bed by waves occur in different seasons. In such cases, the initial deposition may take place near the river but be removed, partially or wholly, by wave-induced resuspension a few months later. Depending on the amount of time that elapses between initial deposition and eventual resuspension, sediments may undergo varying degrees of consolidation making them more resistant to erosion and more likely to remain at the initial deposition site.

In the fourth dispersal stage, the river sediments reach their 'final' resting place and the rate of net accumulation exceeds the rate of erosion and resuspension. It is the accumulated products of this final stage that leave the most lasting geologic record. In the case of the Amazon delta, Pb-210 analyses of cores (Kuehl *et al.* 1986) indicate that on century timescales, roughly half the river's sediment load is accumulating on the mid-shelf (depth 30–50 m) at an average rate of $10\,cm\,yr^{-1}$. The remainder of the sediment is sequestered within the subaerial delta. In contrast, the mild energy regime of the Mississippi Delta, together with a rather rapid rate of tectonic subsidence, has permitted the formation of thick accumulations at the mouths of prograding distributaries on the centuries timescale. On a longer timescale, episodic delta lobe switching has yielded multiple thick and elongate accumulations distributed along the Louisiana coast.

As deltas prograde seaward over a continental shelf and spread laterally along a coast, a subaerial delta plain usually surmounts the underlying subaqueous platform. Although this subaerial surface, which includes the intertidal environments, is typically quite thin vertically, at least in comparison to subaqueous deposits, it is this component that supports most human activities and with which people are most familiar. The suites of geomorphologic features that distinguish any particular deltaic surface are, as for the subaqueous delta, products of the coastal process regime that moulded the delta as well as of the regional climate and human land use practices. Other factors include the degree to which a delta is undergoing submergence because of rising sea level, tectonic subsidence or both. Wright (1985) describes some of the most common subaerial features.

References

Coleman, J.M. and Wright, L.D. (1975) Modern river deltas: variability of processes and sand bodies, in M.L. Broussard (ed.) *Deltas: Models for Exploration*, Houston Geological Society, 99–149.

DeMaster, D.J., McKee, B.A., Nittrouer, C.A., Qian, J. and Cheng, G. (1985) Rates of sediment accumulation and particle reworking based on radiochemical measurements from continental shelf deposits in the East China Sea, *Continental Shelf Research* 4, 153–158.

Geyer, W.R., Hill, P.S., Milligan, T. and Traykovski, P. (2000) The structure of the Eel River plume during floods, *Continental Shelf Research* 20, 2,067–2,093.

Kolb, C.R. and Van Lopik, J.R. (1966) Depositional environments of Mississippi River deltaic, southeastern Louisiana, in M.L. Shirley (ed.) *Deltas in their Geological Framework*, Houston Geological Society, 17–61.

Kuehl, S.A., DeMaster, D.J. and Nittrouer, C.A. (1986) Nature of sediment accumulation on the Amazon continental shelf, *Continental Shelf Research* 6, 209–226.

Milliman, J.D. and Meade, R.H. (1983) World-wide delivery of river sediment to the oceans, *Journal of Geology* 91, 1–21.

Milliman, J.D. and Syvitski, J.P.M. (1992) Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers, *Journal of Geology* 100, 525–544.

Ogston, A.S., Cacchione, D.A., Sternberg, R.W. and Kineke, G.C. (2000) Observations of storm and river flood-driven sediment transport on the northern California continental shelf, *Continental Shelf Research* 20, 2,141–2,162.

Wright, L.D. (1985) River deltas, in R.A. Davis (ed.) *Coastal Sedimentary Environments*, 1–76, New York: Springer-Verlag.
Wright, L.D. and Nittrouer, C.A. (1995) Dispersal of river sediments in coastal seas: six contrasting cases, *Estuaries* 18, 494–508.
Wright, L.D., Wiseman, W.J., Yang, Z.-S., Bornhold, B.D., Keller, G.H., Prior, D.B. and Suhayda, J.M. (1990) Processes of marine dispersal and deposition of suspended silts off the modern mouth of the Huanghe (Yellow River), *Continental Shelf Research* 10, 1–40.

L.D. WRIGHT

RIVER PLUME

A plume is a vertically or horizontally moving, rising or expanding fluid body, such as the contrails of a fighter jet, emissions from a stack, or river discharge into a lake. Under the strong influence of gravity, rivers can enter a lake or ocean as a fully turbulent jet (Baines and Chu 1996), such as from discharge across steep riverbeds (>0.5°), or under the influence of a flood wave. A river will reduce its velocity at its mouth, if it is moving too fast, by undergoing a hydraulic jump and thickening its flow (Bursik 1995). Many rivers discharge more slowly. The river's momentum and the hydraulic head at the river mouth carry the plume up to hundreds of kilometres into the lake or ocean, depending on the size and power of the river (Syvitski *et al.* 1998).

The plume's behaviour is dependent on the density contrast between the river water and the standing water. Compared to most lake water, the contrast in effluent density is small and controlled by the river's suspended load. Ocean water has a higher density, and the plumes often flow buoyantly on the surface (hypopycnal: Plate 99). The pathway that a hypopycnal plume will take, depends on a variety of factors:

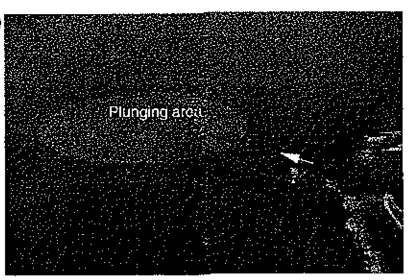
1 Angle between the river course at the entry point and the coastline;
2 Strength and direction of the coastal current;
3 Wind direction and its influence on local upwelling or downwelling conditions;
4 Mixing (tidal) energy near the river mouth; and
5 latitude of the river mouth and thus the strength of the Coriolis effect.

Often there are strong interactions between these factors. For example if the angle of entry is in the direction of the Coriolis effect (i.e. move to the right in the northern hemisphere), then the plume will likely form a coast-hugging plume. Otherwise the plume will detach from the coast.

While hypopycnal plumes may form when river water enters a freshwater lake, they are just as likely to form hyperpycnal plumes (Plate 99). Sometimes referred to as underflows or turbidity currents, these dense flows penetrate the lake under the influence of gravity, remaining in contact with the lake floor (Kassem and Imran 2001). If a hyperpycnal plume accelerates, additional sediment can be resuspended into the flow from the lake floor. As the hyperpycnal plume spreads and thickens due to the entrainment of ambient fluid, velocity is reduced and sediment is deposited. Hyperpycnal plumes rarely occur in rivers that discharge to the ocean (Wright *et al.* 1986). Globally, only a few dozen rivers generate hyperpycnal plumes on an annual basis, and most rivers would see such events happen once every hundred or so years, if at all (Mulder and Syvitski 1995).

Another major difference in comparing sediment plumes flowing into oceans and lakes is in the dynamics of particle settling. In freshwater environments, finer particles settle out of the plume slowly and individually. In ocean environments, river particles quickly flocculate and settle out rather quickly (Syvitski *et al.* 1995). Flocculation is the process that sees particles come into contact with one another and stick, wherein the new larger particles (flocs) have greatly enhanced settling velocities.

River plumes may also enter a lake or ocean at depth: tidewater glaciers directly discharge their stream water at or near the base of the ice front.

(a)


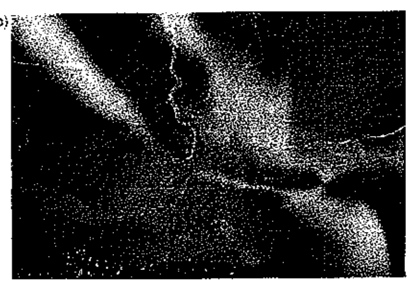
(b)


Plate 99 (a) A hyperpycnal plume forming seaward of Skeiðarársandur (Iceland) (1996 photograph by M.T. Gumundsson and F. Pálsson). The surface plume disappears at the plunging point and subsequently flows seaward along the seafloor. (b) SeaWiFs image showing hypopycnal plumes emanating from the Mississippi River

References

Baines, W.D. and Chu, V.H. (1996) Jets and Plumes, in V.P. Singh and W.H. Hager (eds) *Environmental Hydraulics*, 7–61, Netherlands: Kluwer Academic.
Bursik, M. (1995) Theory of the sedimentation of suspended particles from fluvial plumes, *Sedimentology* 42, 831–838.
Kassem, A. and Imran, J. (2001) Simulation of turbid underflow generated by the plunging of a river, *Geology* 29, 655–658.
Mulder, T. and Syvitski, J.P.M. (1995) Turbidity currents generated at river mouths during exceptional discharges to the world oceans, *Journal of Geology* 103, 285–299.
Syvitski, J.P.M., Asprey, K.W. and LeBlanc, K.W.G. (1995) In-situ characteristics of particles settling within a deep-water estuary, *Deep-Sea Research II* 42, 223–256.
Syvitski, J.P.M., Nicholson, M., Skene, K. and Morehead, M.D. (1998) PLUME1.1: Deposition of sediment from a fluvial plume, *Computers and Geosciences* 24, 159–171.
Wright, L.D., Yang, Z.-S., Bornhold, B.D., Keller, G.H., Prior, D.B. and Wiseman, W.J. (1986) Hyperpycnal plumes and plume fronts over the Huanghe (Yellow River) delta front, *Geo-Marine Letters* 6, 97–105.

JAMES SYVITSKI

RIVER RESTORATION

River restoration is a term used to describe a wide range of approaches aimed at improving the environmental quality of engineered river systems (see CHANNELIZATION; DAM; STREAM RESTORATION). The objective may be to recreate the river's natural forms and processes (sometimes referred to as 'naturalization'), although the assumption that nature can be created has been criticized. Restoration has also been described as 'nudging nature'. This reflects the fact that, to date, much restoration work has been focused on lower energy streams which have less ability to recover naturally following river channelization and therefore require active intervention (see Brookes 1987, for a discussion of the recovery of channel sinuosity on straightened rivers in Denmark). Furthermore, improvements in urban rivers, sometimes undertaken for aesthetic reasons, have also been referred to as 'restoration'. And some restoration schemes may also involve the development of a resource, such as a riverside wetland, that did not previously exist at the site. 'Creation' may therefore be a more appropriate term in these situations. Consequently, there is no simple definition of river restoration. However, Brookes and Shields (1996: 4) make the following useful distinction between enhancement, rehabilitation and restoration.

Enhancement they define to be 'any improvement in environmental quality'. This would include, for example, the increased diversity of marginal river vegetation achieved following works to raise flood banks on the River Torne (UK). The enhancement works comprised bank re-profiling to create narrow wetland shelves (berms), shallow bays, channel margins of varying shape and depth and linear still ponds from borrow pits on the floodplain (Clarke and Wharton 2000). The opportunity for river enhancement, which was achieved at negligible cost, arose because contractors changed the usual practice of importing materials and obtained the spoil for the flood banks from the channel margins and the floodplain. A large number of case studies describing river enhancement techniques are also illustrated in *The New Rivers and Wildlife Handbook* (RSPB *et al.* 1994) and there is much scope for these small-scale river improvements.

Rehabilitation, as defined by Brookes and Shields (1996: 4), is 'the partial return to a pre-disturbance structure or function' (see Brookes and Shields 1996 for examples from northern Europe and the USA). An example from the UK of a small-scale rehabilitation project is the Redhill Brook. This is typical of many lowland streams in England which have undergone rehabilitation since the mid-1980s. A 100 m reach had been artificially straightened and a further realignment was planned in 1991 as part of a floodplain development project. However, in issuing a land drainage consent for this work the National Rivers Authority required the realigned section to be designed to reflect the characteristics of a natural lowland stream. This included creating a channel with varying channel cross sections incorporating pools, riffles and point bars. Sediment was also reinstated which would not erode in a bankfull flood event (Brookes and Shields 1996: 246–247). Much more extensive rehabilitation has been undertaken on the rivers Brede, Cole and Skerne comprising a joint Danish and British EU–LIFE demonstration project (see Holmes and Nielsen 1998 for information on the background to the project; and Kronvang *et al.* 1998 for details on the restoration of the channel morphology and hydrology).

Finally, 'restoration' in its strictest sense is the term employed by Brookes and Shields (1996: 4) to describe 'the complete return to a pre-disturbance structure or function'. There are, however, several constraints to full river restoration. First, there is likely to be disagreement about the most appropriate pre-disturbance state. Practically, there is a need to establish whether the baseline for restoration should be set immediately before the most recent channelization works, before the first evidence for channel modification or at some point in between. Second, there is the related problem of establishing pre-disturbance data. Rarely are data comprehensive and accurate enough for reconstruction to be fully informed. In this context, Tapsell (1995) has asked 'what are we restoring to?' and Graf (1996, 2001) has discussed the issue of what is natural and how closely restored systems can approximate natural conditions. And third, the desirability of river restoration must be questioned. If sustainable river management is the aim, then the river must be considered within its catchment context, with river forms and processes able to respond to controlling factors such as flow regimes and sediment transport rates, that in turn respond to changing drainage basin conditions. A river restored to some pre-disturbance state is unlikely to be in balance with present conditions. Erskine *et al.* (1999) also document how restoration of the pre-dam situation on the Snowy River (Australia) was neither possible nor desirable because the conditions below the dam had stabilized themselves to a new regime.

Thus, the term rehabilitation is more appropriate in reflecting the reality of river restoration. In the UK, the River Restoration Centre, a non-profit making organization working to restore and enhance rivers, views restoration as a visionary target 'of pristine rivers that are wholly returned to an undisturbed state requiring no management' (Holmes 1998: 139) while accepting rehabilitation as a more practical alternative.

The involvement of all stakeholders, including the participation of the public, is a key element in the success of river restoration schemes by helping to engender a sense of ownership by the local community. In the EU–LIFE demonstration project on the River Skerne (UK) a community liaison officer was a vital member of the team working with the experts and local residents to ensure effective public consultation and dialogue (Holmes and Nielsen 1998). And Waley (2000) describes the socio-cultural value of river restoration in Japan and how science and ethics combine in restoration programmes.

The restoration of rivers may be driven by many factors, including environmental, economic and political. Attempts to restore the geomorphology, hydrology, water quality and ecology of rivers may arise from a desire to redress the environmental impacts of past engineering schemes (see CHANNELIZATION; DAM). Thus most river restoration activity is concentrated in developed countries which have a long history of river engineering. And in the UK, improvements to the physical habitat and ecology have been shown to be the main drivers behind restoration initiatives. The removal of dams which no longer generate hydro-electric power at a competitive rate and the restoration of a more natural flow regime and river environment has been reported in the USA (Graf 1996). This can have benefits for wildlife and generate income from the recreational use of the river (e.g. fishing and canoeing). Changes in environmental legislation also have a significant influence on river restoration. For example, in Denmark the 1982 Watercourse Act provides powers for safeguarding the physical environment of streams by focusing on ecologically acceptable maintenance practices, and incorporates special provisions for stream restoration and the potential for financial support of such activities. Across Europe, the European Union Water Framework Directive (2000/60/EC) is providing further impetus to river restoration by requiring member states to protect, enhance and restore all bodies of surface water not designated as artificial or heavily modified.

Brookes (1988: 217–237) and Wharton (2000) describe procedures for restoring channel capacity, river-bed sediments, cross-sectional form and pattern. Clearly, however, these components should not be viewed separately in the process of restoring a river's geomorphology. Based on research in Sweden, Petersen *et al.* (1992) advocate a 'building block approach' for restoring river environments in a number of stages. By combining different elements such as the construction of pools and riffles, the re-meandering of reaches and the creation of buffer strips and wetlands, the design and implementation of the restoration scheme can be tailored to specific sites. Brookes and Shields (1996) have also published guiding principles on river restoration and the UK River Restoration Centre has produced a second edition of its *Manual of River Restoration Techniques* (RRC 2002). By maintaining a database on completed projects and river restoration practitioners and researchers, the RRC also plays a pivotal role in disseminating information on river restoration and forms part of the European Centre for River Restoration (ECRR).

As more river restoration projects are undertaken it becomes increasingly important to appraise these schemes so that failures as well as successes can be documented and evaluated (Kondolf 1998). Importantly, it should be recognized that river restoration can have impacts on the fluvial system similar to those reported for channelization. Information from immediate post-project appraisal and longer term monitoring will help to develop the science of restoration. Specifically, there is a need to improve predictions of river channel sensitivity to change and to incorporate this understanding in integrated catchment management planning. There is also a need for further evidence on the link between the restoration of geomorphology and physical habitat and subsequent improvements to river ecology. And the appraisal of river restoration schemes in terms of their management and implementation will help inform the development of future policy and practice.

References

Brookes, A. (1987) The distribution and management of channelized streams in Denmark, *Regulated Rivers* 1, 3–16.

——(1988) *Channelized Rivers: Perspectives for Environmental Management*, Chichester: Wiley.

Brookes, A. and Shields, F.D. Jr (eds) (1996) *River Channel Restoration, Guiding Principles for Sustainable Projects*, Chichester: Wiley.

Clarke, S.J. and Wharton, G. (2000) An investigation of marginal habitat and macrophyte community enhancement on the River Torne, UK, *Regulated Rivers: Research and Management* 16, 225–244.

Erskine, W.D., Terrazzolo, N. and Warner, R.F. (1999) River rehabilitation from the hydrogeomorphic impacts of a large hydro-electric power project: Snowy River, Australia, *Regulated Rivers: Research and Management* 15, 3–24.

Graf, W.L. (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.

——(2001) Damage control: restoring the physical integrity of America's rivers, *Annals of the Association of American Geographers* 91, 1–27.

Holmes, N.T.H. (1998) The river restoration project and its demonstration sites, in L.C. De Waal, A.R.G. Large and P.M. Wade (eds) *Rehabilitation of Rivers: Principles and Implementation*, 133–148, Chichester: Wiley.

Holmes, N.T.H. and Nielsen, M.B. (1998) Restoration of the rivers Brede, Cole and Skerne: a joint Danish and British EU-LIFE demonstration project, I – Setting up and delivery of the project, *Aquatic Conservation: Marine and Freshwater Ecosystems* 8, 185–196.

Kondolf, G.M. (1998) Lessons learned from river restoration projects in California, *Aquatic Conservation: Marine and Freshwater Ecosystems* 8, 39–52.

Kronvang, B., Svendsen, L.M., Brookes, A., Fisher, K., Moller, B., Ottosen, O., Newson, M. and Sear, D. (1998) Restoration of the rivers Brede, Cole and Skerne: a joint Danish and British EU-LIFE demonstration project, III – Channel morphology, hydrodynamics and transport of sediment and nutrients, *Aquatic Conservation: Marine and Freshwater Ecosystems* 8, 209–222.

Petersen, R.C., Petersen, L.B.-M. and Lacoursiere, J. (1992) A building-block model for stream restoration, in P.J. Boon, P. Calow and G.E. Petts (eds) *River Conservation and Management*, 293–309, Chichester: Wiley.

RRC (2002) *Manual of River Restoration Techniques*, 2nd edition, Silsoe, Bedfordshire, UK: River Restoration Centre.

RSPB, NRA and RSNC (1994) *The New Rivers and Wildlife Handbook*, The Lodge, Sandy, Bedfordshire, UK: Royal Society for the Protection of Birds.

Tapsell, S.M. (1995) River restoration: what are we restoring to? A case study of the Ravensbourne River, London, *Landscape Research* 20, 98–111.

Waley, P. (2000) Following the flow of Japan's river culture, *Japan Forum* 12, 199–217.
Wharton, G. (2000) New developments in managing river environments, in A. Kent (ed.) *Reflective Practice in Geography Teaching*, 26–36, London: Paul Chapman.

SEE ALSO: anthropogeomorphology

GERALDENE WHARTON

ROCHE MOUTONNÉE

Roches moutonnées are asymmetric bedrock bumps or hills with one side ice-moulded and the other side steepened and often cliffed. They are widespread features in formerly glaciated hard-rock terrain and often to be found in clusters or fields. The name was first introduced by de Saussure (1786), based on a fancied resemblance to the wavy wigs of that period, which were called moutonnées after the mutton fat used to hold them in place. The term encompasses a wide range of feature sizes. Typically, roches moutonnées vary from 1 to 50 m in height and a few metres to hundreds of metres in length, but Sugden *et al.* (1992), for instance, describe large roches moutonnées hills in eastern Scotland with lee side cliffs up to 160 m high.

The morphology of roches moutonnées seems to reflect the contrast between ABRASION on the smoothed up-ice side and plucking on the lee side (see GLACIAL EROSION).

Abrasion acting on the stoss side is marked by STRIATIONS at a variety of scales together with polished facets and crescentic fractures on more gently sloping surfaces. As the glacier moves against the upstream side of an obstruction the ice overburden pressure increases. The basal ice may reach the PRESSURE MELTING POINT and partially melt, causing the glacier to slide. The direction of basal ice flow in this position is pointed towards the bed and the embedded clasts are dragged over the bedrock with some force, effectively abrading it.

On the lee side of the obstruction the ice overburden pressure is lower than average, encouraging the formation of a subglacial water-filled cavity. The presence of a cavity together with fluctuations of the water pressure within it strongly promotes the process of glacial plucking. Sugden *et al.* (1992) showed that block removal starts at the furthest point down ice in the cavity and from there extends successively up ice, thereby transforming the lee side of the bedrock bump into a staircase cliff. The detailed morphology of the plucked surface is also influenced by the properties of the parent bedrock, since plucking is encouraged by a favourable oriented system of joints.

Some roches moutonnées do appear to be only slightly modified preglacial hills, but in many areas initial bedrock eminences were clearly sculptured and reshaped by differential glacial erosion. Between these end-members there is likely to be a continuum of forms with varying degree of inherited topography. If the quarried and rough lee side as a distinctive feature of a roche moutonnée is little developed, it may also be difficult to distinguish roches moutonnées from whalebacks or rock DRUMLINS. Whalebacks are elongated and approximately symmetrical bedrock bumps whereas rock drumlins are asymmetrical with a steep upstream side and a gently inclined downstream side. Both are smoothed and rounded on all sides.

Reference

Sugden, D.E., Glasser, N. and Clapperton, C.M. (1992) Evolution of large roches moutonnées, *Geografiska Annaler* 74A, 253–264.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.

CHRISTINE EMBLETON-HAMANN

ROCK COATING

About 15 per cent of the Earth's landscape consists of bare rock surfaces. Yet the common phrase 'bare rock' is truly a misnomer, because paper-thin accretions coat almost all of these rock surfaces in all terrestrial environments. Studies on the physical and chemical characteristics, origin, geography and utility of these deposits has spawned over 3,000 scientific papers. Plate 100 illustrates a few examples.

Alexander von Humboldt (1812) initiated the scholarly study of rock coatings by studying the composition, origin, spatial distribution and environmental relations of coatings such as those found along tropical rivers. In the past two centuries, researchers have documented hundreds of different types of rock coatings found within the fourteen major categories listed in Table 39.

The three most common rock coatings are rock varnish (see DESERT VARNISH), silica glaze and iron films. Silica glazes occur in warm deserts, cold

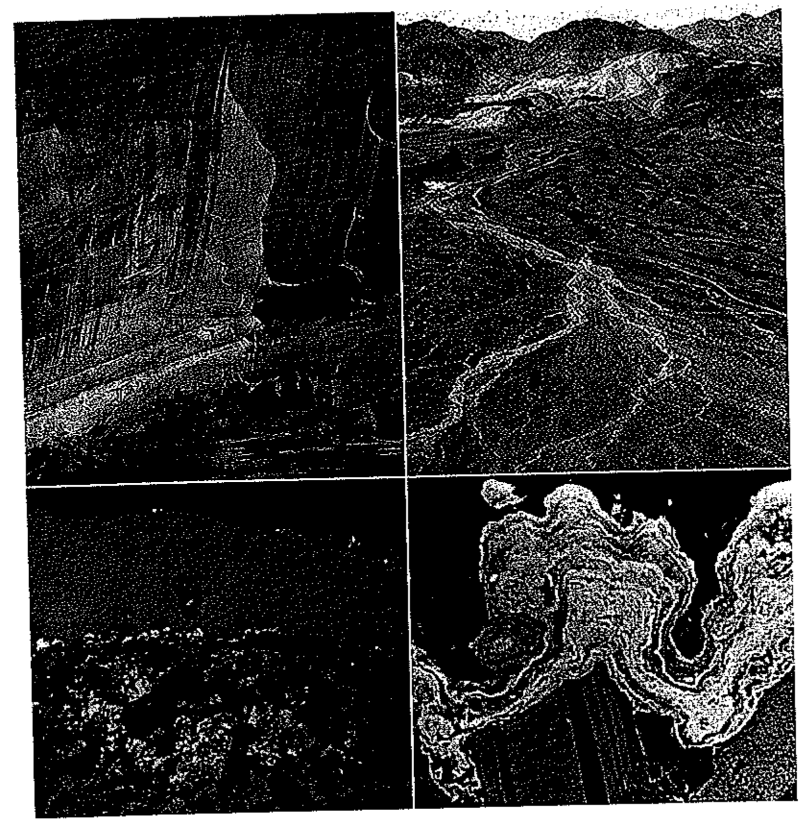


Plate 100 Upper left: a vertical face at Canyon De Chelly, USA, is streaked with heavy metal skins, iron films, lithobiotic coatings, oxalate crusts, rock varnish and silica glaze. Upper right: alluvial fan in Death Valley deposits the same light-coloured rock types in active streams. But over time, rocks in abandoned stream courses are darkened by desert varnish. Lower row: the electron microscope images (backscatter detector) illustrate that rock coatings are external accretions, exemplified by an oxalate crust in the lower left image that is about 0.5 mm thick and desert varnish on the lower right that is about 0.1 mm thick

deserts like Antarctica, on dry tropical islands, along tropical rivers, mid-latitude humid temperate settings, and various archaeological settings. Silica glazes probably precipitate from soluble Al-Si complexes $[Al(OSi(OH)_3)]^{2+}]$ that are released from the weathering of clay minerals. Rust-coloured iron films display a wide variety of characteristics in very different climates and microenvironments. For example, rocks in the Dry Valleys of Antarctica host iron hydroxides that both form a micron-scale accretion and a weathering rind (see RIND, WEATHERING) over a millimetre thick. In a very different setting, iron oxyhydroxides impregnate rocks in arctic streams (Dixon *et al.* 2002).

Geomorphologists have long used intuition to interpret rock coatings and their relationship to the geomorphic setting. For example, some have believed that PEDIMENTS are fossil landforms, in part because the presence of rock coatings must

Table 39 Major categories of rock coatings

General type	Description	Related terms
Carbonate skin	Coating composed primarily of carbonate, usually calcium carbonate, but could be combined with magnesium or other cations	Caliche, calcrete, patina, travertine, carbonate skin, dolocrete, dolomite
Case hardening agents	Addition of cementing agent to rock matrix material; the agent may be manganese, sulphate, carbonate, silica, iron, oxalate, organisms or anthropogenic	Sometimes called a particular type of rock coating
Dust film	Light powder of clay- and silt-sized particles attached to rough surfaces and in rock fractures	Gesetz der Wüstenbildung; clay skins; clay films; soiling
Heavy metal skins	Coatings of iron, manganese, copper, zinc, nickel, mercury, lead and other heavy metals on rocks in natural and human-altered settings	Described by chemical composition of the film
Iron film	Composed primarily of iron oxides or oxyhydroxides; unlike orange rock varnish because it does not have clay as a major constituent	Ground patina, ferric oxide coating, red staining, ferric hydroxides, iron staining, iron-rich rock varnish, red-brown coating
Lithobiontic coatings	Organic remains form the rock coating, e.g. lichens, moss, fungi, cyanobacteria, algae	Organic mat, biofilms,
Nitrate crust	Potassium and calcium nitrate coatings on rocks, often in caves and rock shelters in limestone areas	Saltpetre; nitre; icing
Oxalate crust	Mostly calcium oxalate and silica with variable concentrations of magnesium, aluminium, potassium, phosphorus, sulphur, barium and manganese. [illegible] found forming near or with lichens. Usua[illegible]ark in colour, but can be as light as ivory	Oxalate patina, lichen-produced crusts, patina, scialbatura
Phosphate skin	Various phosp[illegible] minerals (e.g. iron phosphates or [illegible]) with clays and so[illegible]	Organ[illegible] film[illegible]
Pigment	Human-manufactured material placed on rock surfaces by people	Pictogr[illegible] sometimes described by the nature of the material
Rock varnish	Clay minerals, Mn and Fe oxides, and minor and trace elements; colour ranges from orange to black produced by variable concentrations of different manganese and iron oxides	Desert varnish, desert lacquer, patina, manteau protecteaur, Wüstenlack, Schutzrinden, cataract films
Salt crust	The precipitation of sodium chloride on rock surfaces	Halite crust, efflorescence, salcrete

Table 39 Continued

General type	Description	Related terms
Silica glaze	Usually clear white to orange shiny lustre, but can be darker in appearance, composed primarily of amorphous silica and aluminium, but often with iron	Desert glaze, turtle-skin patina, siliceous crusts, silica-alumina coating, silica skins
Sulphate crust	Composed of the superposition of sulphates (e.g. barite, gypsum) on rocks; not gypsum crusts that are sedimentary deposits	Gypsum crusts; sulphate skin

infer long-term stability. Others have guessed at the ages of such features as flooding events on ALLUVIAL FANS, based on an intuitive feeling about the appearance of rock coatings (see gradual darkening of alluvial fan surfaces in the Plate 100). The complexities associated with formative processes have made rock coatings extraordinarily difficult to use as geomorphological tools to indicate either age or infer palaeoclimate. Rock coatings will be getting increased attention in future years as they are identified on Mars and as planetary scientists attempt to use rock coatings to infer Martian geomorphic processes (Kraft and Greeley 2000).

Rock coatings have applied significance in a variety of contexts. Heavy metal skins assist in identifying metal pollution (Dong *et al.* 2002). Some believe that artificial rock coatings have potential to aid in the conservation of priceless stone monuments (Borgia *et al.* 2001). Construction and development in desert regions contrasts bright uncoated rocks and darker natural rock coatings; the desire to live in natural-appearing settings leads to the application of artificial rock coatings to mimic natural colouration (Henniger 1995). Rock coatings, called patina in archaeology, are also used in the study of surface artefacts, rock paintings and rock engravings.

References

Borgia, G.C., Bortolotti, V., Casmaiti, M., Cerri, F., Fantazzini, P. and Piacenti, F. (2001) Performance evolution of hydrophobic treatments for stone conservation investigated by MRI, *Magnetic Resonance Imaging* 19, 513–516.

Dixon, J.C., Thorn, C.E., Darmody, R.G. and Campbell, S.W. (2002) Weathering rinds and rock coatings from an Arctic alpine environment, northern Scandinavia, *Geological Society of America Bulletin* 114, 226–238.

Dong, D., Hua, X. and Zhonghua, L. (2002) Lead adsorption to metal oxides and organic material of freshwater surface coatings determined using a novel selective extraction method, *Environmental Pollution* 119, 317–321.

Henniger, J. (1995) Fooling mother nature with Permeon artificial desert varnish, *Rocky Mountain Construction* 76(8), 48–52.

Kraft, M.D. and Greeley, R. (2000) Rock coatings and aeolian abrasion on Mars: application to the Pathfinder landing site, *Journal of Geophysical Research – Planets* 105, 15,107–15,116.

von Humboldt, A. (1812) *Personal Narrative of Travels to the Equinoctial Regions of America During the Years 1799–1804 V. II*, translated and ed. T. Ross in 1907, London: George Bell & Sons.

Further reading

Dorn, R.I. (1998) *Rock Coatings*, Amsterdam: Elsevier.

SEE ALSO: alluvial fan; desert pavement; desert varnish; pediment

RONALD I. DORN

ROCK CONTROL

Rock control in geomorphology is defined as the influences of differences in earth materials on the development of landforms. Earth materials that form the Earth's surface or landforms are simply called landform materials, and include rocks, weathered materials and soil. The concept of *rock control* was first proposed explicitly and argued passionately by Yatsu (1966) and then expanded by him to a concept of *landform material science* in 1971. Yatsu stressed that to understand the formation of landforms, it is necessary to quantitatively evaluate the behaviours of landform materials in terms of their physical, mechanical, chemical and mineralogical properties in relation to the geomorphological processes concerned. His severe criticism of geomorphology is based on the fact that geologic structure and lithology have only been used qualitatively to explain the development of erosional landforms since the birth of modern geomorphology.

Typical examples often described in textbooks on geomorphology (e.g. Thornbury 1954; Sparks 1971) as structural landforms or landforms resulting from rock control include cuestas, hogbacks, mesas, structural benches, dyke ridges, knickpoints, karst and inversion topography. These landforms are relatively higher or steeper than their surroundings, and are generally composed of a relatively *resistant* or *hard* rock (e.g. sandstone, limestone, lava) that adjoins the relatively *less resistant* or *weak* rocks (e.g. mudstone, shale, tuff). However, rock control is not as simple as the vague terms *resistant* and *less resistant imply*. This is because the resistance and behaviour of landform materials vary markedly with geomorphological process and geomorphic setting.

For instance, the rocky coast of Arasaki, southwest of Tokyo, Japan, is underlain by steeply dipping, alternating beds of mudstone and scoria tuff. Differential erosion between the two rocks varies with altitude (Figure 134). On the vegetation-free sea cliffs behind the uplifted wave-cut benches, mudstone forms ridges and tuff forms shallow furrows. On the benches, mudstone forms the furrows and tuff forms the ridges, producing a washboard-like relief. On the shallow offshore sea bottom, there is no differential erosion. The mudstone is mechanically about two times as strong as the tuff. However, it is well jointed and forms fragments about 1 cm in size due to wet–dry slaking, whereas the tuff does not. The explanation for this differential erosion is that (1) both rocks are eroded at rates proportional to their mechanical strengths on the sea cliff above tidal zone, (2) the fragments of mudstone produced by wet–dry slaking are rapidly washed away by waves in the tidal zone, and (3) wave abrasion offshore erodes both rocks at the same rate (see Suzuki 2002).

Thus, the behaviour of landform materials generally does not merely depend on their geological structure and lithology, but also strongly depends on their physical and mechanical properties. This is because even lithologically similar rocks have wide ranges of physical and mechanical properties,

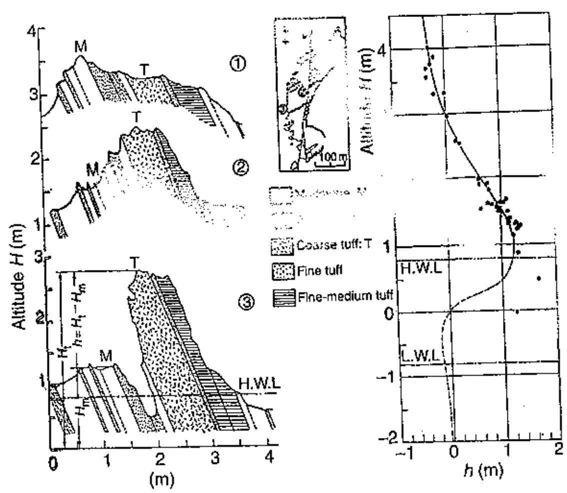


Figure 134 Change in relative relief between a tuff bed (T) and a mudstone bed (M) with altitude (H) on the Arasaki coast, Japan. *Left*: three geologic sections that are different in the altitude (H_m) of the mudstone surface (M). *Right*: relationship between relative relief (h) and H_m

reflecting their origin and history, such as diagenesis, tectonic deformation, unloading and weathering. Further, weathering results in changes in rock properties and hence is of importance as a preparatory process for erosion and mass movements.

Properties of landform materials are grouped into two major categories: geological properties and rock (material) properties. Geological properties are those described in terms of geology, and include lithology (such as grain size, mineral composition and texture), chemical composition, geological structure (such as stratification, joints, faults, unconformities, and their dip and strike), occurrence of rock masses (such as lava flow, dyke, batholith, etc.), weathering grades, and so on. Rock (material) properties, on the other hand, are those described in terms of physics and engineering (particularly rock and soil mechanics), and which can be further subdivided into physical properties and mechanical properties. Physical properties are the intrinsic characteristics (such as density, porosity and pore-size distribution) that do not depend on applied forces. Mechanical properties are the behaviours and responses of rocks against forces acting upon them, and hence vary with the kind of forces, the conditions of the rocks, such as water content, and the test methods used. Mechanical properties include strength (compressive, tensile, shear and bending strengths), hardness (e.g. abrasion, impact and scratch), deformation properties (e.g. deformation modulus, adhesive forces and internal friction), dynamic properties (elastic wave velocities), thermal properties (e.g. thermal conductivity and thermal expansion coefficient), permeability (e.g. permeability coefficient and infiltration capacity), behaviour in relation to water (such as swelling, slaking and solution) and so on.

These physical and mechanical properties are determined by the measurements of the rock mass in the field and for test pieces in the laboratory using precise instruments and equipment. Some practical test methods have also been applied to evaluate the mechanical properties of rocks, including standard penetration tests (N-value), rebound hardness with a Schmidt rock hammer, cone penetration hardness, needle penetration hardness for the rock mass, and rock quality designation for drilling cores.

Rock control problems are addressed by looking for the important rock properties in each process and quantitatively evaluating the roles of the properties with respect to the process. In the context of Yatsu's argument, therefore, physical, mechanical and chemical properties of landform materials have been intensively measured in both field and laboratory and in field and laboratory experiments since the 1970s, particularly by Japanese geomorphologists. Based on the measurements and experiments, much persuasive substantiation has been found for the modes and rates of various erosional processes and landforms (Suzuki 2002). Notable examples include formative processes along rocky coasts (Sunamura 1992), wind abrasion, lateral planation, slope evolution, hillslope morphology, valley development and some minor landforms such as tafoni. Processes and rates of bedrock weathering have also been studied actively in both field and laboratory.

Landform evolution is controlled not only by rock properties and geological properties but also by many other variables, such as geomorphological setting (initial landform), climate, geomorphic agents from which the various forces derive, and elapsed time. The research on rock control problems mentioned above, therefore, has been directed toward developing quantitative models of geomorphological processes that are capable of predicting types and rates of landform development. The models have been expressed as geomorphological equations including the geomorphic quantities concerned and the main controlling variables, i.e. site- and time-independent process equations with dimensionless constants. To develop the models, it is indispensable to study the rock control problems all over the world, because landforms are never changed unless the landform materials are moved. Thus, research on the rock control problems will be one of the core fields in process geomorphology in the twenty-first century.

References

Sparks, B.W. (1971) *Rocks and Relief*, London: Longman.

Sunamura, T. (1992) *Geomorphology of Rocky Coasts*, Chichester: Wiley.

Suzuki, T. (2002) Rock control in geomorphological processes: research history in Japan and perspective, *Transactions, Japanese Geomorphological Union* 23, 161–199.

Thornbury, W.D. (1954) *Principles of Geomorphology*, New York: Wiley.

Yatsu, E. (1966) *Rock Control in Geomorphology*, Tokyo: Sozosha.

——(1971) Landform material science – rock control in geomorphology, in E. Yatsu, F.A. Dahms, A. Falconer, A.J. Ward and J.S. Wolfe (eds) *Research Method in*

Geomorphology (Proceedings of the 1st Guelph Symposium on Geomorphology, 1969), 49–56, Ontario: Science Research Associates.

Further reading

Selby, M.J. (1993) *Hillslope Materials and Processes*, Oxford: Oxford University Press.
Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

SEE ALSO: rock mass strength; weathering

TAKASUKE SUZUKI

ROCK AND EARTH PINNACLE AND PILLAR

Within areas built of poorly consolidated sediments, subject to intensive linear erosion, sheet wash and susceptible to weathering, bedrock may be sculpted into groups of weirdly shaped erosional residuals in the form of pinnacles, pillars and cones. They are relatively common in semi-arid areas, where scarce vegetation provides little protection against surface erosion, hence pinnacles are typical components of BADLANDS. Steep slopes underlain by erodible sediments, for example of newly deposited MORAINEs, may also support pinnacle assemblages.

Two types of sediments yield to this type of erosional relief in particular, tills and pyroclastic deposits. Some tills and other glacial deposits contain boulders 'floating' in an otherwise fine-grained material. After exposure, boulders will protect an underlying softer mass against erosion, whereas the surrounding unprotected material will be eroded away, leaving the boulder-capped part standing as a residual pillar. Later, the boulder cap will provide a shield against the direct destructive impact of rain and the pillar may increase in height as long as the cap remains in place. Once the boulder falls from the top of the pinnacle, the column built of soft rock will rapidly be destroyed. Classic localities of this type of earth pinnacles have been described from the Tyrol in the Alps.

In pyroclastic deposits, volcanic bombs within softer tuff play the similar protective role as boulders in tills do. In Cappadocia, central Turkey, bomb-capped pyramids reach up to 20 m high. In other cases, caps are provided by remnants of a welded tuff horizon overlying thicker and softer strata beneath.

Not all rock and earth pinnacles have a protective cap, and there are other reasons why they remain as isolated residuals. In the semi-arid badlands of Cappadocia, surfaces of tuff cones isolated by fluvial and sheet wash erosion are subject to case hardening, and it is the crust which protects the cones from further destruction. Owing to the presence of the crust, the earth pinnacles of Cappadocia could have grown up to 15–20 m high and were found stable enough for churches, hermitages and cave dwellings to be dug into them in early Christian times.

Tufa (see TUFA AND TRAVERTINE) deposits may form curiously shaped pinnacles too, but in these cases pinnacles are constructional, and not erosional features. At Mono Lake, California, and Searles Lake, California, tufa pyramids and pillars up to 15 m high formed around underwater springs and were exposed at the surface, when lake levels were lowered.

Plate 101 A group of rock pinnacles in Cappadocia, central Turkey. Remnants of resistant welded tuff act as a cap to the underlying softer sediment

SEE ALSO: hoodoo

PIOTR MIGOŃ

ROCK GLACIER

Rock glaciers (German: *Blockgletscher, Blockstrom*; French: *Glacier rocheux*; Russian: *Kammeni gletscher*; Spanish: *glaciar rocoso*) occur in most alpine-mountainous regions and are distinct tongues of rock rubble which flow slowly downhill. Most features are elongate and are generally distinct from blockstreams (see BLOCKFIELD AND BLOCKSTREAM) which occur on very low angle slopes.

The substantial literature on these features is complex and often confusing, being hindered by difficulties of *in situ* investigation. Markedly different viewpoints have been taken to explain their origin, dynamic behaviour and environmental significance (Potter *et al.* 1998). It is important to avoid explicit designation of an origin so they are best defined by their morphology and appearance; a simple morphological definition, after Washburn (1979), is: 'a tongue-like body of angular boulders that resembles a small glacier, which generally occurs in high mountains and usually has ridges, furrows and sometimes lobes on its surface with a steep front or snout at the angle of repose'.

Plate 102 An active rock glacier (maximum velocity about $0.25\,\mathrm{m\,a^{-1}}$) in northern Iceland with a small corrie glacier at its head. There is a gradation from a thin cover of debris to much thicker debris (*c.*1 m) near the snout. The lateral margins are distinct from the sides of the valley and the longitudinal furrow is conspicuous. The whole rock glacier is about 800 m long and estimated to have formed in the past 200–300 years

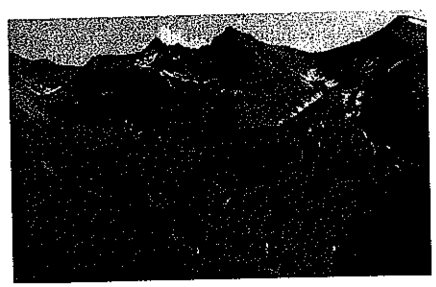


Plate 103 Merging rock glaciers, Wrangell Mountains, Alaska. These are typical rock glacier forms, emerging from corries now containing little or no ice. That there is probably very little forward movement of the wrinkled and furrowed surfaces is indicated by the vegetated surface

Distribution maps and reviews can be found in Whalley and Martin (1992) and Barsch (1996). They may even occur on Mars (Whalley and Azizi 2003). Although originally thought to be confined to continental areas and to give way to glacier bodies in more maritime regions examples in the latter have been found. They were first recognized in North America and Greenland (Martin and Whalley 1987; Barsch 1996).

The surface velocity is generally $<1\,\mathrm{m\,a^{-1}}$, although some with velocities $>5\,\mathrm{m^{-1}}$ have been described (Gorbunov *et al.* 1992). If no movement can be detected they are generally referred to as 'relict' and, if highly vegetated with subdued features, as 'fossil' and are recognized by morphology alone. However, 'inactive' rock glaciers are sometimes recognized where creep rates are very low and may even have trees growing on them. The steep fronts (snouts) may advance over other features; valley floors, moraines and lakes. These characteristics, variable over time, make it difficult to show that they result from one origin or relate to a single set of environmental conditions.

Rock glaciers are generally about 1 km long but many shorter examples exist and some may be up to 3 km long; width is generally a few hundred metres. Typically, they have their heads in corries (see CIRQUE, GLACIAL) in which there may be a small glacier, although this may not always be visible. The elongate forms are regarded as rock glaciers proper, but have also been called 'tongue-shaped', 'valley floor' or 'debris rock glaciers'. Some forms may be 'spatulate' where they spread over a main valley floor. Rock glaciers are usually separated from valley sides by 'lateral furrows'. The term 'rock glacier' has also been used for features which are broader than long and which typically have their upper sections against cliffs or scree rather than emanating from corries. This form has been called a 'valley side rock glacier', 'lobate rock glacier' or 'talus rock glacier'. It has been suggested that the latter features are best termed 'protalus lobes' rather than rock glaciers because of the differences in form and location (Hamilton and Whalley 1995).

Flow-related features are commonly seen as ridges and furrows on the surface although it is not known if these extend to any depth. Some rock glaciers have mainly transverse ridges, especially near the snout, others have a predominantly longitudinal ridge pattern. Such flow features have been related to flow regime; compressing or extending (Whalley and Azizi 2003). Many rock glaciers have distinct longitudinal furrows and some show pools or small lakes developed in flatter areas ('thermokarst ponds').

There are four main theories of rock glacier origin. One suggests that they are glacially derived with a veneer of weathered rock debris (a few cm to >1 m thick) over a thin (<50 m) glacier ice core. The debris protects a thin body of ice which flows only slowly. This may be termed the 'glacial model'. It has been suggested that rock glaciers are nothing more than debris-covered glaciers. However, the subdued dynamic behaviour of rock glaciers indicates that ice volumes are small and that SUPRAGLACIAL debris has been supplied via the surface of the small glacier. Debris-covered glaciers gain debris in their lower reaches by ablation of ice releasing englacially transported debris but there is probably a transition between the two. What gives rise to rock glaciers is the relative abundance of debris to active glacier ice.

The 'permafrost model' explains the slow movement as creep of ice dispersed within weathered rock debris (derived from SCREE) and that a glacial derivation is not necessary to explain flow. It does require PERMAFROST conditions (mean annual air temperature <−1.5 °C) for the formation and continued creep of the ice. The ice needs to be above 'saturation', i.e. more than fills void spaces, or as ice lenses, for creep to be efficient. Ice-cemented debris (at or below saturation) colder than the PRESSURE MELTING POINT of ice is mechanically stronger than ice and will not flow unless at high shear stresses (high surface slope and/or thickness).

The third model suggests that some rock glaciers (or protalus lobes) are formed by sudden, perhaps catastrophic, failure of scree slopes (Johnson 1974) or as a single catastrophic rock avalanche (Bergsturz or STURZSTROM). This view is not widely held although there is evidence that *some* rock glacier forms might have been constructed in this way to produce topography similar to a rock glacier. Where the features are old there might be confusion with a fossil rock glacier.

A fourth model, a variant of the first (glacier-derived) and third (catastrophic), is that a rockfall covered a small or decaying glacier. The thin debris cover would thus insulate the thin glacier core but suddenly rather than progressively.

Plate 104 Complex protalus lobes, Alpes Maritimes, southern France. The inner ridges look a little like protalus ramparts and the feature lies below a cliffed area which probably had a thin glacier or large snowpatch at its foot. These features differ in form to rock glaciers per se

Rock glaciers have been used as indicators of permafrost (past permafrost for relict features) but only if the permafrost model is valid. This may be considered as being a 'zonal' model. The glacial model is thus 'azonal' as the contributing glacier may occur whether or not permafrost is present (Washburn 1979).

The origin and flow mechanism of rock glaciers is frequently attributed exclusively to creeping permafrost (Barsch 1996). Although traces of glacier ice have been seen in some rock glaciers, permafrost conditions were considered to be the only way in which the features could exist. Observations of glacier ice down the length of some rock glaciers have now been established and thus show that at least some rock glaciers have glacier ice cores. It is possible that modern dating and isotopic techniques will allow ice from such rock glaciers to provide a climatic record. The full implications for recognizing climate change through rock glaciers still needs investigation.

Geophysical measurements (seismic, gravity, resistivity and ground penetrating radar) have been used to investigate the structure of rock glaciers. Resistivity has been used to differentiate between the mode of ice formation. It is claimed that high resistivity (>10 MΩm) is indicative of glacier ice but that rather lower values are typical of ice of permafrost origin. This has been disputed by some authors who claim that the high resistivity is not typical of the small glaciers which provide glacier cores because such ice is contaminated by dust which lowers the value. The difficulty is of linking geophysical signals with an appropriate

mixture (ice and debris) model (Whalley and Azizi 1994). The complexity is enhanced because there may be grades of mixture, from permafrost to glacier ice core, in one feature and is particularly significant near rock glacier snouts. This ambiguity of origin also suggests that using rock glaciers to identify past conditions may be difficult.

'Protalus lobes' are related to rock glaciers where permafrost conditions may be required to preserve ice but where a glacier is unlikely to have formed. These are distinctive enough to be given a separate name. PROTALUS RAMPARTs are long, rather narrow, ridges below cliffs and are thought to have a snow-bank (nival) origin. Suggestions have been made that they might be precursors to rock glaciers of permafrost origin (Barsch 1996).

References

Barsch, D. (1996) *Rock Glaciers*, Berlin: Springer.

Gorbunov, A.P., Titkov, S.N. and Polyakov, V.G. (1992) Dynamics of rock glaciers of the northern Tien Shan and the Djungar Ala Tau, Kazahkstan, *Permafrost and Periglacial Processes* 3, 29–39.

Hamilton, S.J. and Whalley, W.B. (1995) Rock glacier nomenclature: a re-assessment, *Geomorphology* 14, 73–80.

Johnson, P.G. (1974) Mass movement of ablation complexes and their relationship to rock glaciers, *Geografiska Annaler* 56A, 93–101.

Martin, H.E. and Whalley, W.B. (1987) Rock glaciers: Part 1: rock glacier morphology: classification and distribution, *Progress in Physical Geography* 11, 260–282.

Potter, N. Jr, Steig, E.J., Clark, D.H., Speece, M.A., Clark, G.M. and Updike, A.B. (1998) Galena Creek rock glacier revisited – new observations on an old controversy, *Geografiska Annaler* 80A, 251–265.

Washburn, A.L. (1979) *Geocryology: A Survey of Periglacial Processes and Environments*, London: Arnold.

Whalley, W.B. and Azizi, F. (1994) Models of flow of rock glaciers: analysis, critique and a possible test, *Permafrost and Periglacial Processes* 5, 37–51.

Whalley, W.B. and Azizi, F. (2003) Rock glaciers and protalus landforms: analogous forms and ice sources on Earth and Mars, *Journal of Geophysical Research, Planets* 108(E4), art.no. 8,032.

Whalley, W.B. and Martin, H.E. (1992) Rock glaciers: Part II: models and mechanisms, *Progress in Physical Geography* 16, 127–186.

BRIAN WHALLEY

ROCK MASS STRENGTH

Rock mass strength (RMS) refers to the specific properties of the rock mass that control its strength and subsequent rock slope stability. Importantly, it allows the prediction of the stable inclination of natural rock slopes, as well as the recognition of strength EQUILIBRIUM SLOPEs in the landscape adjusted to the prevailing SUBAERIAL processes. The standard method of RMS determination applied in geomorphology was developed by Selby (1980, 1982) and has been extensively tested over the past twenty years as a reliable method for the assessment of rock slope stability. This classification scheme is a modification of RMS classifications developed for engineering purposes (e.g. Bieniawski 1979) which have been extensively used to aid excavation design for tunnels, slopes and foundations. However, these classifications often do not incorporate a quantitative assessment of the influence of a reduction in rock mass strength due to WEATHERING. Furthermore, the engineering classifications contain different definitions of the rating classes so that the results derived from the various methods are not directly comparable. The method of Selby (1980), and modified by Moon (1984), explicitly and quantitatively incorporates and weights the influence of weathering on the strength of the rock mass in the field through evaluation of intact rock strength, estimation of state of rock weathering, joint spacing, continuity and infilling. Since weathered rock is the norm, the scheme developed by Selby (1980) is a more appropriate measure of the RMS of a natural rock slope than those developed for engineering purposes.

Although geomorphologists have long recognized that rock slope failure often occurs along discontinuities such as joints, bedding planes and faults (see FAULT AND FAULT SCARP) rather than through intact rock, logistical difficulties and frequent inability to access the appropriate equipment for the laboratory and field assessment of rock strength has often meant that studies of the strength of the rock mass tended to be qualitative in nature. It is in this context that the development of the rock mass strength classification has had important consequences for our understanding of the morphology and evolution of rock slopes as it provides a basis for understanding the features of the rock mass that provide resistance to weathering and EROSION, as well as the maintenance of slope stability. The only equipment required is a SCHMIDT HAMMER, tape measure and inclinometer, and it can be applied to any rock mass where there is enough exposure to allow measurement of the rock JOINTING (usually at least 10 m^2). If the slope contains more than one morphological element, it must be subdivided into zones with similar RMS properties, with the RMS assessment undertaken within each slope element.

The rock mass strength classification system developed by Selby (1980, 1982) was based on an examination of rock slopes in Antarctica and New Zealand, and has subsequently been applied in a range of settings (e.g. Augustinus 1992; Moon and Selby 1983; Allison and Goudie 1990). Slopes adjusted to their RMS (strength equilibrium slopes) are common in nature, and the recognition of over steepened slopes that have been undercut by erosion, as well as structurally controlled rock benches of lower slope angle and RICHTER DENUDATION SLOPES, indicates its utility in resistance-form studies. The RMS classification involves measurement of a range of properties of the rock mass: (1) Schmidt hammer impact as a measure of the strength of the intact rock; (2) state of rock weathering; (3) jointing characteristics of the rock mass: spacing of rock joints, joint width, joint continuity, joint infilling and orientation of the dominant joint set; and (4) water seepage from the rock face (Table 40). Since not all these parameters are of equal importance in controlling rock mass strength, each of these factors is weighted and given a rating value according to their perceived influence on stability of the rock slope using the scheme given in Selby (1980) or as modified by Moon (1984). However, the usefulness of the further subdivision of the classification proposed by Moon has been questioned, so that the simpler scheme of Selby (1980) is preferred. The sums of the individual weightings for the rock mass being evaluated is its RMS rating. A maximum value of 100 applies and the range is divided into five classes (Table 40). The higher the RMS rating, the higher mass strength and the steeper the slope inclination that can be sustained.

Table 40 Geomorphic rock mass strength classification and ratings

Criteria	*(1)* Very strong	*(2)* Strong	*(3)* Moderate	*(4)* Weak	*(5)* Very weak
Intact rock# strength	100–60	60–50	50–40	40–35	35–10
Rating	*20*	*18*	*14*	*10*	*5*
Weathering	Unweathered	Slightly weathered	Moderately weathered	Highly weathered	Completely weathered
Rating	*10*	*9*	*7*	*5*	*3*
Joint spacing	>3 m	3–1 m	1–0.3 m	0.3–0.05 m	<0.05 m
Rating	*30*	*28*	*21*	*15*	*8*
Joint orientation	>30° into slope	<30° into slope	Horizontal and vertical	<30° out of slope	<30° out of slope
Rating	*20*	*18*	*14*	*9*	*5*
Joint width	<0.1 mm	0.1–1 mm	1–5 mm	5–20 mm	>20 mm
Rating	*7*	*6*	*5*	*4*	*2*
Joint continuity	None continuous	Few continuous	Continuous, no infill	Continuous, thin infill	Continuous, thick infill
Rating	*7*	*6*	*5*	*4*	*1*
Groundwater outflow	None	Trace	Slight	Moderate	Great
Rating	*6*	*5*	*4*	*3*	*1*
Total rating	100–91	90–71	70–51	50–26	<26

Source: Modified from Selby (1980), and Moon (1984)
Note: # N-type Schmidt hammer rebound values

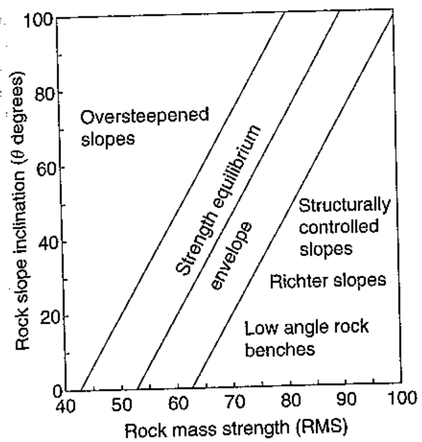


Figure 135 Plot of slope gradient and rock mass strength, with strength equilibrium envelope (after Moon 1984; Abrahams and Parsons 1987)

The graphical representation of the RMS classification involves plotting the total RMS rating against the slope inclination at each site (Figure 135). Note that the slope (θ) vs RMS graph is accompanied by an equilibrium line which relates the RMS rating to the stable slope angle, as defined by numerous measurements of slopes assessed to be in a stable equilibrium condition (Selby 1980, 1982). Superimposed on this plot is the RMS envelope as modified from that of Selby (1980) by Moon (1984), and further refined by Abrahams and Parsons (1987). Within this envelope there is a 95 per cent probability that the slopes are in strength equilibrium (Figure 135). Abrahams and Parsons (1987) re-evaluated the published RMS data for strength equilibrium slopes and produced a more statistically rigorous relationship between slope inclination and RMS. Using this plot and envelope, predictions of stable slope angles can be achieved, and it is possible to identify equilibrium or non-equilibrium slopes, with the latter either oversteepened or low angle, structurally controlled or Richter denudation slopes (Figure 135).

Strength equilibrium slopes have inclinations in balance with their RMS and are not controlled by other exo- or endogenic processes such as structural or tectonic factors. These slopes also require considerable time for this balance to develop (>10,000 years) so that young slopes are often not in strength equilibrium (Selby 1987). Nevertheless, many slopes have an inclination adjusted to their RMS, and oversteepened slopes undercut by processes such as GLACIAL EROSION can be easily recognized, although Augustinus (1995) demonstrated that equilibrium and structurally controlled slopes are more common in youthful, tectonically active mountains with deeply incised glacial valleys. Furthermore, the widespread recognition of strength equilibrium rock slopes suggests that many of them probably retreat whilst preserving strength equilibrium (Moon and Selby 1983; Selby 1987). Consequently, a change in the slope angle during retreat can occur where RMS changes as a consequence of progressive WEATHERING or rapid rupture of the rock mass as a consequence of external factors such as earthquake shaking. The tendency for slopes to equilibrate rapidly as a consequence of a change in RMS means that oversteepened slopes will have a short life span (on a geological timescale) before they evolve towards strength equilibrium forms as soon as fractures open, increase in continuity, widen or rotate.

The importance of rock mass control in geomorphology is exemplified by the application of the rock mass strength classification to the development of an understanding of rock slope form evolution and stability. However, rock mass strength and its resistance to EROSION processes may also be crucial to understanding the evolution of erosional landforms. For example, the development of features such as glacial valley longitudinal profiles (as well as the glacial valley cross-profile forms) will be dependent on the RMS of the rock being eroded as well as the EROSIVITY of the processes, since the intact rock strength, orientation of the rock joints and their spacing will influence the rock mass resistance to glacial erosion processes such as plucking. Clearly, in this situation the rock mass properties that control the stability and morphology of slopes will not be applicable to quantifying rock resistance to erosion and would require redefinition for this purpose.

References

Abrahams, A.D. and Parsons, A.J. (1987) Identification of strength equilibrium rock slopes: further statistical considerations, *Earth Surface Processes and Landforms* 12, 631–635.

Allison, R.J. and Goudie, A.S. (1990) The form of rock slopes in tropical limestone and their associations with rock mass strength, *Zeitscrift für Geomorphologie* 34, 129–148.
Augustinus, P.C. (1992) The influence of rock mass strength on glacial valley cross-profile morphometry: a case study from the Southern Alps, New Zealand, *Earth Surface Processes and Landforms* 17, 39–51.
——(1995) Rock mass strength and the stability of some glacial valley slopes, *Zeitschrift für Geomorphologie* 39, 55–68.
Bieniawski, Z.T. (1979) *Engineering Rock Mass Classifications*, New York: Wiley.
Moon, B.P. (1984) Refinement of a technique for determining rock mass strength for geomorphological purposes, *Earth Surface Processes and Landforms* 9, 189–193.
Moon, B.P. and Selby, M.J. (1983) Rock mass strength and scarp forms in Southern Africa, *Geografiska Annaler* 65A, 135–145.
Selby, M.J. (1980) Rock mass strength classification for geomorphic purposes, *Zeitschrift für Geomorphologie* 24, 31–51.
——(1982) Controls on the stability and inclinations of hill slopes formed on hard rock, *Earth Surface Processes and Landforms* 7, 449–467.
——(1987) Rock Slopes, in M.G. Anderson and K.S. Richards (eds) *Slope Stability*, 475–504, Chichester: Wiley.

Further reading

Selby, M.J. (1993) *Hillslope Materials and Processes*, 2nd edition, Chapter 6, Oxford: Oxford University Press.

PAUL AUGUSTINUS

ROCKFALL

Rockfall is the free or bounding fall of rock debris down steep slopes. Rockfalls vary from individual pebbles to catastrophic failures of several million cubic metres (STURZSTROM; rock avalanches). Smaller rockfalls ($<10^1$–$10^2\,m^{-3}$) are the primary process associated with the formation of SCREE (talus) slopes and may be classified into two types (Rapp 1960). Massive vertical cliffs are dominated by primary rockfalls where detachment is followed by direct transfer to the scree below. These are triggered mainly by pressure release or FREEZE–THAW CYCLE activity. However, debris may accumulate on irregularities in the cliff (benches, gullies, etc.) and subsequently be dislodged by other rockfalls, snow avalanches (see AVALANCHE, SNOW), surface flows, animals, etc. These secondary rockfalls have different magnitude–frequency characteristics than primary rockfalls. Triggering mechanisms for rockfalls are inferred from inventory studies that compare the pattern of rockfalls with simultaneous observations of temperature and precipitation. Most investigators identify diurnal maxima during times of solar illumination and seasonal maxima in spring and fall (Luckman 1976; Gardner 1980).

References

Gardner, J.S. (1980) Frequency, magnitude and spatial distribution of mountain rockfalls and rockslides in the Highwood Pass area, Alberta, Canada, in D.R. Coates and J. Vitek (eds) *Thresholds in Geomorphology*, London: Allen and Unwin.
Luckman, B.H. (1976) Rockfalls and rockfall inventory data: some observations from Surprise Valley, Jasper National Park, Canada, *Earth Surface Processes* 1, 287–298.
Rapp, A. (1960) Talus slopes and mountain walls at Templefjorden, Spitzbergen, *Norsk Polarinstitutt Skrifter* 119.

SEE ALSO: geomorphological hazard; hillslope, process; mass movement; pressure release; unloading

BRIAN LUCKMAN

ROCKPOOL

Rockpools (synonymous with tidepool, pool) can broadly be defined as depressions in eulittoral and supralittoral rocky SHORE PLATFORMS which store surface water and form as a result of dissection of rock material by a combination of chemical, physical and/or biological means. It is generally accepted that the presence or creation of an initial depression enables the commencement of a positive feedback loop – where surface water storage provides a suitable environment where weathering and erosion processes widen and deepen pits in rock surfaces to develop rockpools (Elston 1917).

Rockpool development can be divided into three phases: (1) pool initiation, (2) pool widening and deepening and (3) coalescing of smaller pools. Pool initiation is thought to be largely controlled by geological conditions such as rock hardness (with softer rocks such as sandstone and limestone being more prone to erosion), joint planes, irregular bedding and concretions which provide an initial depression from which rockpools gradually develop. Pool deepening and widening are often caused by a suite of biological, chemical and physical processes. Biological processes include bioerosion by boring and/or grazing species such as polychaete worms, sea urchins and limpets. Dissolution is often caused by biochemical activity when respiration increases the CO_2 concentration in the pool. Chemical weathering also occurs due to evaporation and drying of seawater. The dominant physical erosion process is scouring and abrasion of pools by harder rock debris being moved or rotated by the action of waves. The third phase of pool formation is caused by the continual erosion of narrow walls separating adjacent pools. This leads to the coalescing of smaller pools into large irregular or elliptical pool forms and, in some instances, secondary, inset pools develop as another line of stratification is eroded in the base of pools.

Rockpools vary in size from small features a few centimetres in diameter to large, irregular forms which are up to 6 m in diameter and range from 0.1 to 2 m in depth. Two main types of rock pools have been defined in geomorphological literature: solution pools and pot-holes. Solution pools (synonymous with shallow pools) are typically defined as shallow, flat-bottomed depressions found on gently sloping shore platforms (Sunamura 1992). While solution pools have greater width than depth and vary in shape, pot-holes are typically cylindrical or bowl-shaped depressions that have more similar depth to width ratios. Pot-holes are thought to form primarily by abrasion. This classification is quite narrow in scope as rockpools typically form by a combination of biological, chemical and physical means rather than being dominated by one process over another.

References

Elston, E.D. (1917) Potholes: their variety, origin and significance, *Scientific Monthly* 5, 554–567.
Sunamura, T. (1992) *Geomorphology of Rocky Coasts*, Chichester: Wiley.

Further reading

Emery, K.O. (1946) Marine solution basins, *Journal of Geology* 54, 209–228.
Emery, K.O. and Kuhn, G.G. (1980) Erosion of rock shores at La Jolla, California, *Marine Geology* 37, 197–208.
Wentworth, C.K. (1944) Potholes, pits and pans: subaerial and marine, *Journal of Geology* 52, 117–130.

LARISSA NAYLOR

ROCKY DESERTIFICATION

Rocky DESERTIFICATION is the process that leads to KARST lands being turned into stony ecological deserts. It is a consequence of devegetation followed by intensive agriculture and extreme soil erosion. Soils on karsts are usually thin, because limestones are often composed of at least 90 per cent calcium carbonate and so there is little insoluble residue from their solution that can form the mineral basis of a soil. In the humid subtropical karst of China, it has been calculated that 0.25–0.85 million years are required to form 1 m of soil. Where thick soils are found on karst, it is usually because they are formed of foreign materials transported from beyond the boundary of the karst, such as loess, alluvium, volcanic ash or glacial drift. But even thick soils can be stripped from karst, although it takes longer.

It is well known in any environment that deforestation leads to accelerated SOIL EROSION. In karst this is exacerbated because of soil loss down countless voids opened by corrosion into caves, where it has a major deleterious impact on subterranean biota. It is eventually evacuated from cave systems via underground streams that discharge at springs, but lowered water quality results. The free draining nature of karst therefore contributes to the loss of its soil if the fragile hold accomplished by plant cover is disturbed.

In parts of the Mediterranean basin, the rocky nature of karst is so characteristic that it has come to be considered natural, rather than a consequence of millennia of human impact (Gams *et al.* 1993). The word *karst* itself can be traced back to pre-Indo-European origins, where it stems from *karra* meaning stone. But the landscapes involved were originally forested and have been cleared, tilled and overgrazed. The first evidence of forest clearance around the northern Mediterranean was in the Neolithic about 6,000 years ago. This continued through Greek, Roman and more recent times, and as population increased lands were subdivided and grazing became more intensive. This relentless impact contributed to the stripping of the hills; so that naked karst now seems the norm. But recent political and land use changes in Slovenia have seen rural migration and abandonment of farms, followed by natural regrowth and spread of forests, indicating that recovery is possible if human pressure is reduced.

A similar sequence of events has occurred in China (Yuan 1995), especially in Guizhou Province, where removal of subtropical monsoon forest and intense population pressure in recent centuries has seen the denudation of karstic hillsides. The expansion of rocky desertification in the area was at a rate of 933 square kilometres per year during the 1980s. Similar problems though of smaller scale are encountered in deforested and intensively farmed karstlands of the Gunung Sewu of Java and in parts of Central America and the Caribbean.

References

Gams, I., Nicod, J., Sauro, U., Julian, E. and Anthony, U. (1993) Environmental change and human impacts on the Mediterranean karsts of France, Italy and the Dinaric region, in P.W. Williams (ed.) *Karst Terrains: Environmental Changes and Human Impact*, Cremlingen-Destedt, *Catena Supplement* 25, 59–98.

Yuan Daoxian (1995) Rock desertification in the subtropical karst of South China, in P.W. Williams (ed.) Tropical and subtropical karst, *Zeitschrift für Geomorphologie, Supplementband* 108, 81–90.

PAUL W. WILLIAMS

ROUGHNESS

The term 'roughness' refers to the roughness of a channel bed, which is an important component of the overall resistance to water flow along the channel. Water flows along a channel under the influence largely of two forces: the downslope component of its own weight (which acts to propel it along the channel) and the resistance of the channel (which acts to hold it back). If the resistance is low, then a given flow has a high velocity and a low depth. If the resistance is high, the same flow has a low velocity and a high depth. Quantification of flow resistance is thus fundamental to the calculation of flow conditions in a channel.

Several relationships linking flow resistance, velocity and depth have been in use for a century or more, each quantifying the resistance with a single coefficient. They are the Darcy–Weisbach equation:

$$U = (8 g R S_f / f)^{1/2}$$

the Manning equation (in SI units):

$$U = R^{2/3} S_f^{1/2} / n$$

and the Chézy equation:

$$U = c (R S_f)^{1/2}$$

where U is mean flow velocity, R is hydraulic radius (flow cross-sectional area/channel wetted perimeter), S_f is friction slope (a measure of energy loss often approximated by water surface slope), g is acceleration due to gravity and f, n and c are respectively, the Darcy–Weisbach, Manning and Chézy coefficients. The central problem in flow resistance is therefore evaluation of the coefficient.

In a straight channel of uniform slope, uniform cross section and large width/depth ratio with no sediment transport or BEDFORMS, the resistance to flow is determined primarily by the frictional resistance of the bed. This varies with the roughness of the bed, itself dependent on the material of which the bed is composed, e.g. sand or gravel. In a popular approach, fluid mechanics and boundary layer theory are invoked to calculate resistance as a function of the logarithm of relative roughness, defined as the ratio of bed roughness height to flow depth. Roughness height is often evaluated as an equivalent sand grain size or as a selected percentile from the measured size distribution of the bed material. For example

$$(8/f)^{1/2} = 5.62 \log (d/D_{84}) + C$$

where D_{84} is the particle size for which 84 per cent of the particles are finer and C is a coefficient. Because theory cannot yet provide a full quantification of the resistance, the coefficient is determined empirically from experimental data.

Natural channels rarely conform to the ideal conditions described above and additional terms may be required in the resistance equation to account for deviations from these conditions. For example, sand beds develop bedforms such as ripples and dunes which increase the resistance above that of the roughness of the grains alone. Consequently there are a variety of formulae and methods for determining the resistance coefficient. Users should be careful to select a method which is appropriate for the conditions and data availability with which they are concerned.

Further reading

Bathurst, J.C. (1993) Flow resistance through the channel network, in K. Beven and M.J. Kirkby (eds) *Channel Network Hydrology*, 69–98, Chichester: Wiley.

Raudkivi, A.J. (1998) *Loose Boundary Hydraulics*, Rotterdam: Balkema.

JAMES C. BATHURST

RUGGEDNESS

A property of the landscape which describes the complexity of the topography and the roughness of the terrain. More rugged landscapes tend to exhibit a greater amount of complexity, having rough and uneven surfaces. Ruggedness is a naturally qualitative term, though several ruggedness indexes have been proposed (e.g. Riley *et al.* 1999) that provide a quantitative frame. Melton (1958) developed the ruggedness number to describe the ruggedness of land on a drainage-basin scale. This is a dimensionless number calculated from the formula $H / \sqrt{A}$ where H is the vertical relief above fan apex (miles^2), and *A* is basin area (miles^2). In general, the ruggedness number can be as high as 2.0 or 3.0 for a first or second-order basin, and rarely above 1.0 for a third or fourth-order basin.

References

Melton, M.A. (1958) Geometric properties of mature drainage basins and their representation in a E_4 phase space, *Journal of Geology* 66, 35–54.

Riley, S.J., DeGloria, S.D. and Elliot, R. (1999) A terrain ruggedness index that quantifies topographic heterogeneity, *Intermountain Journal of Sciences* 5, 23–27.

STEVE WARD

RUNOFF GENERATION

Runoff generation refers to a suite of processes that produce and route flow from landscape segments to stream channels in response to precipitation (i.e. rainfall and/or snowmelt). Runoff is generated by three different mechanisms: infiltration-excess overland flow (Horton-type), saturation-excess overland flow (Dunne-type) and subsurface stormflow. Infiltration-excess overland flow is overland flow that results from saturation from above (Horton 1933, 1945). This occurs where the water-input rate exceeds the infiltration capacity of the soil long enough for ponding to occur; the excess water then flows quickly over the surface to stream channels (see QUICK FLOW). Once the precipitation volume exceeds the moisture storage capacity of the soil, however, saturation-excess overland flow occurs. First described in detail by Dunne and Black (1970), this mechanism is controlled by the saturated hydraulic conductivity of the soil. The soil becomes saturated from below, either by (1) the presence of an impeding layer which causes a perched water table to develop that may gradually rise to the surface; (2) extension of the capillary fringe to the ground surface; or (3) the presence of a permanent water table at or near the ground surface. Saturation-excess overland flow consists of direct water input to the saturated area plus the return flow contributed by the exfiltration of ground water from upslope.

The third runoff mechanism, subsurface stormflow, consists mainly of the displacement of old pre-event water by new rainwater. This is due to near-stream ground water mounding (the same process that ultimately produces saturation overland flow) and/or via flow from perched saturated zones (saturated wedges). The process also includes throughflow or interflow, which is water that infiltrates into the soil and moves laterally within the soil matrix either in the unsaturated zone between the ground surface and a perched or regional water table or through macropores such as cracks, root and animal holes and pipes (Wilson *et al.* 1990; Jones 1997). The latter reaches the stream channel quickly and differs from other subsurface flow by the rapidity of its response and its relatively large magnitude.

It is widely accepted that Hortonian overland flow is the dominant response mechanism in semi-arid to arid regions, on impermeable zones and in human-disturbed areas. On the other hand, the response mechanism in humid regions is generally saturation-excess overland flow and/or some form of subsurface flow. Saturation-excess overland flow most frequently occurs near stream channels but it also can be generated in hillslope hollows, where subsurface flowlines converge in slope concavities, at concave slope breaks, or where soil layers conducting subsurface flow are locally thin. Subsurface storm flow dominates where soils are deep and permeable, especially under forest cover, and where slopes are steep. All three runoff mechanisms can occur simultaneously in a basin, even during a single water-input event.

References

Dunne, T. and Black, R.D. (1970) Partial area contributions to storm runoff in a small New England watershed, *Water Resources Research* 6, 1,296–1,311.

Horton, R.E. (1933) The role of infiltration in the hydrologic cycle, *Transactions of the American Geophysical Union* 14, 446–460.
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.
Jones, J.A.A. (1997) Pipeflow contributing areas and runoff response, *Hydrological Processes* 11, 35–41.
Wilson, G.V., Jardine, P.M., Luxmoore, R.M. and Jones, J.R. (1990) Hydrology of a forested hillslope during storm events, *Geoderma* 46, 119–138.

SEE ALSO: quick flow

MICHAEL SLATTERY

S

SABKHA

A sabkha is the English form of the Arabic word *sebkha* which means 'salt flat'. Kinsman (1969) defines a sabkha as a surface of deflation down to the level of ground water or the zone of capillary evaporation. Neal (1975) describes a sabkha as a geomorphic surface the level of which is dictated by the water table. Warren (1989) and Briere (2000) describe a sabkha as a marginal marine and continental mudflat where evaporite minerals are forming in the capillary zone above the saline water table. Sabkhas were described subsequently in many other areas of the world, such as the coast of Baja California, Mexico and the coast of Sinai. Equivalent features are the Solonchak salt flats of the Caspian Sea and some kavir depressions of Iran. Certain salt pans in South Africa and playas in the southwestern United States may be similar but true sabkhas are not fed by streams or runoff.

In North America, the words 'playa' and 'salina' have both been applied to sabkha-like areas in the desert. Holm (1960) states that 'playa' is synonymous with 'mamlahah' (inland sabkha). Von Engeln (1942), on the other hand, says that if the percentage of salts in a playa is high enough for a salt crust to form when the flat is dry, it is then called a salina.

If the word sabkha means 'salt flat' then there are both coastal and continental sabkhas. Some coastal sabkhas, such as those along the Trucial Coast, pass laterally into continental sabkhas without any noticeable change in surface morphology on the sabkha (Kinsman 1969). The marine portion of the Abu Dhabi sabkha is characterized by a matrix of marine sediments soaking in largely marine-derived ground water and the continental portion by non-marine sediments and ground water.

Modern marine sabkhas are forming along the coasts of many stable land areas such as the western and southern coasts of the Arabian Gulf, the coasts of Australia and northern Africa. Some sabkhas are slightly above present sea level (0.5–3.0 m) and may have been inherited from short Mid-Holocene eustatic oscillations.

A review of the global distribution of sabkha indicates its extensive presence in the Middle East, including Egypt, Sudan, Libya, Tunisia, Algeria and Ethiopia. Sabkha also exists in India, Australia and southern Africa. Contrary to expectations, sabkha and sabkha-like sediments can occur also in relatively cold climates. Aridity, therefore, seems to play a more important role than hot weather in the formation of sabkha. Figure 136 shows the distribution of sabkha around the world.

Sabkhas are part of a landform sequence that extends from the shoreline, with barrier islands or dunes, through a lagoon then to the sabkha and perhaps into a dune system before truly terrestrial systems are reached. Sabkha surfaces are extremely flat and often extend for long distances. The sabkha depositional sequence can be divided into three units: the subtidal, the intertidal and the supratidal. When the sequence is prograding the three units are superimposed one on top of the other to form a shallowing upward sabkha cycle.

The subtidal unit is usually divided into open marine and restricted marine sedimentation. Restricted marine sedimentation occurs on the landward side of barrier islands. The lagoon sediments contain a diverse biota of many benthic species, molluscan sands occur in the more restricted areas and pelleted carbonate mudstones occur in the more open areas of the

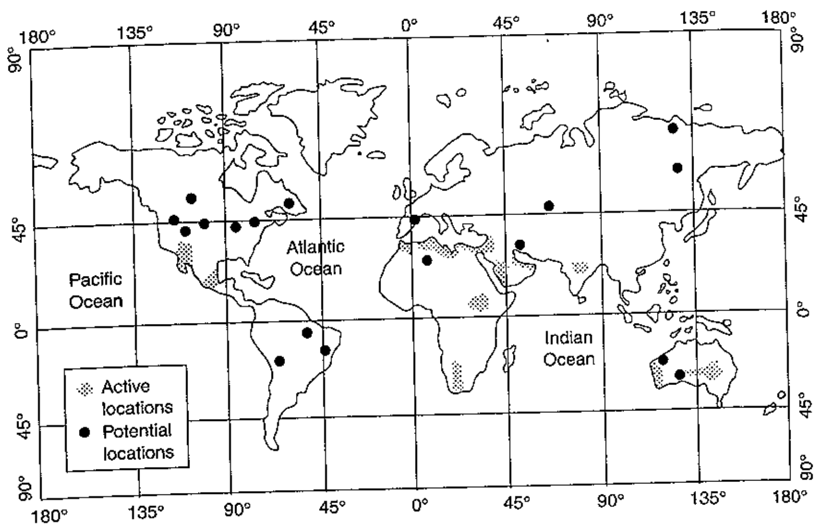


Figure 136 Map of the world showing active and potential sabkha locations (Al-Amoudi 1994)

lagoon. The intertidal zone can be divided into an upper and lower intertidal facies. However, the lower intertidal facies may be the same as the lagoonal facies as described by a number of workers. This facies is dominated by an algal peat composed of the bioturbated remnants of an algal mat. The upper intertidal facies is usually a laminated algal mat often cross-cut by desiccation cracks containing lenticular gypsum crystals. Aragonite, magnesite and dolomite may locally cement the surface sediments (Butler *et al.* 1982). The lower supratidal belt, which is flooded once or twice a month, is characterized by gypsum up to 30 cm thick. Diagenetic nodular anhydrite occurs in the deposits of the middle supratidal zone and there is often a surface crust of ephemeral halite. Such deposits are flooded on less than monthly intervals. In the upper supratidal zone, flooded once every four to five years, the gypsum has been replaced by coalesced nodules of anhydrite.

Sabkhas are broad coastal supratidal and intertidal flats developed along the margins of arid landmasses. Sediments that accumulate on sabkhas include: (1) siliciclastic detritus sediments that are eroded from adjacent land and washed onto the sabkha; (2) offshore deposits of sand and mud that periodic storms wash up and onto the sabkha; and (3) the indigenous sediments of the sabkha itself.

Much of the evaporite sediment produced in sabkhas precipitates as saline ground water seeps into and out of the sabkha (Figure 137). Much of this ground water is seawater that is continually recharged beneath the sabkha, but ground water from the adjacent landmass can also feed the system. Groundwater circulation is driven by capillary action and evaporative pumping. Intermittent flooding by the sea also occurs, and beach ridges can trap a reservoir of additional seawater.

Typical sabkha evaporite minerals are anhydrite, gypsum and dolomite. Much of the anhydrite occurs as irregularly shaped lumps or nodules. These nodules replace altered gypsum crystals originally formed within layers of interbedded carbonate mud or shale. The term chickenwire structure is used to refer to this mixture of elongate, irregular clumps of anhydrite separating thin stringers of carbonate and/or siliciclastic mud (Plate 105a,b). This structure is particularly common in sabkha evaporites.

Cyclicity is common in sabkha evaporite sequences. As deposition proceeds, sabkha deposits

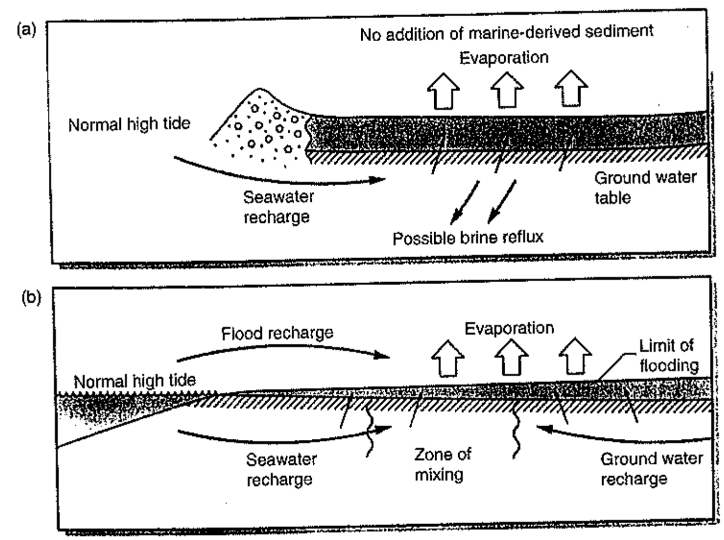


Figure 137 Sabkhas receive water from a variety of sources: (a) sabkha with seawater recharged through the subsurface and with relatively little groundwater influx; (b) sabkha groundwaters are recharged by a mixture of seawater and groundwater, plus seawater flooding from major storms (from Walker 1984)

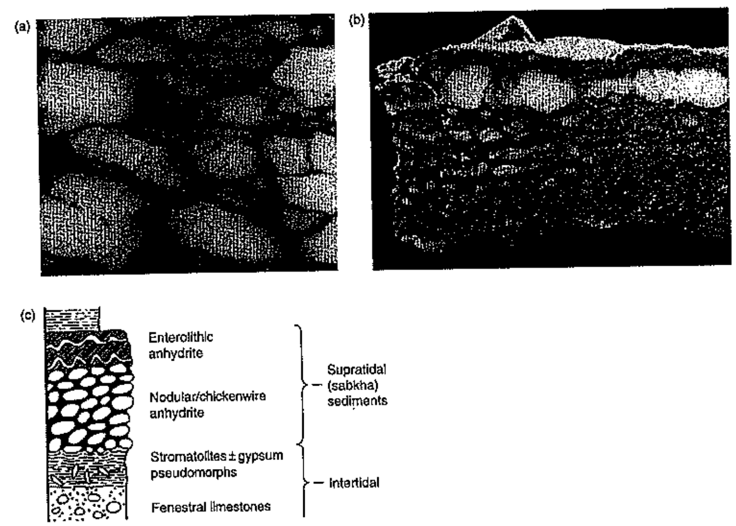


Plate 105 Sabkhas produce a number of distinctive structures: (a) mosaic anhydrite (chickenwire structure) commonly formed when anhydrite nodules coalesce, shown at actual size; (b) nodular gypsum is common just below the surface of the sabkha; (c) typical vertical cycle of sabkha sediments. Such cycles range from several metres to several tens of metres in thickness (after Tucker 1981)

naturally prograde oceanward and eventually lie upon intertidal sediments (STROMATOLITES, gypsum, fenestral birdseye pelleted carbonate mud). These in turn rest on oolitic and bioclastic subtidal carbonate rocks of the subtidal zone (Plate 105c).

The Arabian Gulf coastal sabkhas occur on the southern margin of the Gulf with the best-studied sequences being situated south-east of Abu Dhabi City, where they are now partially covered by urban and industrial development. If an idealized landward transect is traced, it passes in order through offshore open-marine skeletal sediments, belts of oolitic grainstones, belts of lagoonal and/or barrier sequences, and then crosses intertidal algal mats to terminate in a supratidal sabkha sequence (Butler *et al.* 1982). If the intertidal algal mats are included, the sabkhas constitute a zone 10–15 km wide, with an along-strike continuity of more than 150 km. The surface transect from the lagoonal sands and muds up onto the flat plain of sabkha first crosses the dark, black to grey, flat-laminated, algal mats of the intertidal zone (Figure 138).

In some arid areas the surface of the sabkha is encrusted by a veneer of salt and scattered discoidal crystals of gypsum. Dust and sand storms occur through most of the year, which results in the sabkha plain being covered by a layer of drifting quartz sand.

The landward boundary of the supratidal area is characterized by a zone of vegetation and is dominated by the halophyte *Halocnemon Strobilaceum*. The vegetation on the sabkha surface acts as a trap for the moving sand. Most of this drifting sand is usually washed off during storm tides and redistributed on the sabkha surface.

Inland sabkhas develop where water flowing in wadis intermittently floods low-lying depressions (see PAN) to leave behind damp, salt-encrusted sediments. They are also found in depressions where, for one reason or another, the water table reaches the surface.

In inland sabkhas, the salt crust forms as the result of the concentration of salts caused by evaporation of the water. Gypsum crystals are common in the sediments of inland sabkhas. Algae are known, but the algal mats so commonly associated with coastal sabkhas have not been recognized. A coastal sabkha, on the other hand, is characterized by marine flooding and evaporitic conditions. It is a diagenetic environment whose sediments are of continental and adjacent marine origin.

References

Al-Amoudi, O.S.B. (1994) A state-of-the-art report on the geotechnical problems associated with sabkha soils and methods of treatment, *Proceedings of the ASCE-SAS 1st Reg. Conference Exhibition*, Bahrain, 18–20 Sept. 53–77.

Briere, P.R. (2000) Playa, playa lake, sabkha: proposed definitions for old terms, *Journal of Arid Environments* 45, 1–7.

Butler, G.P., Harris, P.M. and Kendall, C.G. St.C. (1982) Recent evaporites from the Abu Dhabi coastal flats, in depositional and diagenetic spectra of evaporites, *A Core Workshop: SEPM Core Workshop* 3, Calgary, 33–64.

Holm, D.A. (1960) Desert geomorphology in the Arabian Peninsula, *Science* 132(3,427), 1,369–1,379.

Kinsman, D.J.J. (1969) Modes of formation, sedimentary associations and diagnostic features of shallow water and supratidal evaporites, *American Association Petroleum Geologists Bulletin* 53, 830–840.

Neal, J.T. (1975) Playa surface features as indicators of environment, in J.T. Neal (ed.) *Playas and Dried Lakes*, 363–380, Benchmark Papers in Geology, Stroudsburg, PA: Dowden, Hutchinson and Ross.

Tucker, M.E. (1981) *Sedimentary Petrology: An Introduction*, 161–173, Oxford: Blackwell Scientific.

Von Engeln, P.D. (1942) *Geomorphology*, New York: Macmillan.

Walker, R.G. (1984) *Facies Models*, 2nd edition, Toronto: Geological Association of Canada.

Warren, J.K. (1989) *Evaporite Sedimentology*, Upper Saddle River, NJ: Prentice-Hall.

SEE ALSO: deflation

ADEEBA E. AL-HURBAN

SACKUNG

A German term describing a type of slope-sagging, gravitational lateral spreading or deep-seated gravitational deformation in mountainous alpine landscapes. Sackungen (plural) display rounded morphology, commonly trend parallel to the contours of the slope, and form a characteristic ridge-top trench in the adjacent valley. They typically display a bulge at the toe

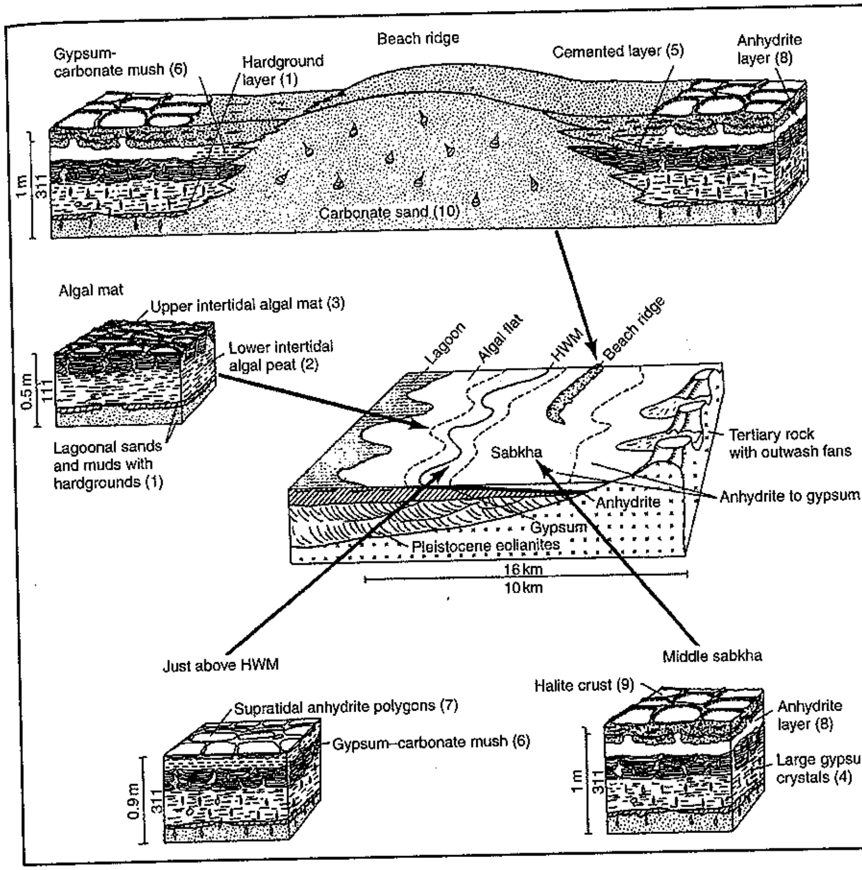


Figure 138 Schematic block diagrams showing sediment and evaporite distribution in Abu Dhabi sabkha, Arabian Gulf. All peripheral diagrams are keyed to central block of sabkha. HWM = high-water mark. (1) Lagoonal carbonate sands and/or muds with carbonate hardgrounds; (2) vaguely laminated lower tidal-flat carbonate-rich algal peat; (3) upper tidal-flat algal mat formed into polygons; (4) large gypsum crystals (many lenticular); (5) cemented carbonate layer; (6) high tidal-flat to supratidal mush of gypsum and carbonate; (7) supratidal anhydrite polygons with wind-blown carbonate and quartz; (8) anhydrite layer replacing gypsum mush and forming diapiric structures; (9) halite crust, formed into compressional polygons; (10) deflated beach ridge of cerithid coquina and carbonate sand. (After Butler *et al.* 1982)

of the slope, known as the 'Talzuschub', as well as tensile or normal faults near their crest, termed 'Bergzerreißung' (Zischinski 1969). Rates of material creep vary significantly (1 mm to several metres), with variations in activity corresponding to changes in precipitation and water table (increased creep activity with higher precipitation). Sackungen are indicative of gravitational spreading (Varnes *et al.* 1989), resulting from low mass strength in the underlying material (a product of high density jointing and faulting) to a substantial depth.

References

Butler, G.P., Kendall, C.G. St. C. and Harris, P.M. (1982) Abu Dhabi Coastal Flats, in C.R. Handford, R.E. Loucks and G.R. Davies (eds) *Depositional and Diagenetic Spectra of Evaporites a Core Workshop*. Society of Economic Palaeontologists Core Workshop, number 3, Calgary, Canada, 33–64.

Varnes, D.J., Radbruch-Hall, D.H. and Savage, W.Z. (1989) Topographic and structural conditions in areas of gravitational spreading of ridges in the western United States, *US Geological Survey Professional Paper* 1496, 28.

Zischinski, U. (1969) Uber sackungen, *Rock Mechanics* 1, 30–52.

Further reading

Savage, W.Z. and Varnes, D.J. (1987) A model for the plastic spreading of steep-sided ridges ('Sackung'), *Bulletin of the International Association for Engineering Geology* 35, 31–36.

SEE ALSO: mass movement

STEVE WARD

SALCRETE

A salty surface crust, primarily composed of sodium chloride, that cements a sand surface as a result of evaporation of moisture and the consequent chemical concentration of dissolved material. The term, which was coined by Yasso (1966), has been mainly used to describe crusts developed through evaporation of sea spray on beaches (e.g. Pye 1980).

References

Pye, K. (1980) Beach salcrete in North Queensland, *Journal of Sedimentary Petrology* 50, 257–261.

Yasso, W.E. (1966) Heavy mineral concentrations and sastrugi-like deflation furrows in a beach salcrete at Rockaway Point, *Journal of Sedimentology Petrology* 36, 836–838.

A.S. GOUDIE

SALT (EVAPORITE) KARST

Evaporites, including common salt (halite) and calcium sulphate (gypsum and anhydrite) are the most soluble of common rocks (see GYPSUM KARST). They are also widespread, being found, for example, in 32 of the 48 contiguous states of the USA (Johnson 1997) in rocks of every geologic system from the Precambrian through to the Quaternary. Karst-creating evaporites are also extensively developed in Canada (Ford 1997), and have been described from many parts of Europe, including the Ripon area of England and the Betic Cordillera and Ebro basins of Spain. Salt dome structures are often affected by solutional processes (see SALT RELATED LANDFORMS).

Evaporite outcrops display a full array of solutional features, including subsidence depressions, collapse breccias, sinkholes, vertical shafts and water-filled chimneys (Last 1993; Calaforra and Pulido-Bosch 1999; Gutierrez-Elorza and Gutierrez-Santollala, 1998). Because of the great solubility of evaporites, rates of karst denudation can be high, and even under the dry conditions of the arid parts of Israel, can approach 500–750 mm 1,000 yr^{-1} (Frumkin 1994). Salt karst processes present a range of engineering problems and hazards (Paukstys *et al.* 1999) and these can be exacerbated as a result of human activities, including mining, ground water abstraction and other hydrological modifications.

References

Calaforra, J.M. and Pulido-Bosch, A. (1999) Gypsum karst features as evidence of diapiric processes in the Betic Cordillera, Southern Spain, *Geomorphology* 29, 251–264.

Ford, D.C. (1997) Principal features of evaporite karst in Canada, *Carbonates and Evaporites* 12, 15–23.

Frumkin, A. (1994) Hydrology and denudation rates of halite karst, *Journal of Hydrology* 16, 171–189.

Gutierrez-Elorza, M. and Gutierrez-Santollala, F. (1998) [illegible]morphology of the Tertiary gypsum formation [illegible] the Ebro Depression (Spain), *Geoderma* [illegible]29.

Johnson, K[illegible] Evaporite [illegible] United [illegible] *Evapo*[illegible]

[illegible] dissolution [illegible] saline lakes of the norther[illegible] eat-Plains [illegible] anada, *Geomorphology* 8, 32[illegible]

Paukstys, B., Cooper, A.H. and Arustiene, J. (1999) Planning for gypsum geohazards in Lithuania and England, *Engineering Geology* 52, 93–103.

A.S. GOUDIE

SALT HEAVE OR HALOTURBATION

A cause of damage to engineering structures in desert areas as a result of the presence of soluble salts (including sodium chloride, magnesium sulphate and sodium sulphate). The process is akin

to frost heave, in that the hydration and crystallization of salts plays a major role, but it is also akin to needle-ice (pipkrake) formation in that salt whiskers may grow vertically. The problem is especially severe when saline ground water approaches the ground surface. Possible techniques to deal with it have been developed (Horta 1985), including brooming, embankments, barriers and the use of thick, impervious surfacings. In the case of some gypsum areas, volume changes associated with solution and recrystallization can produce what are termed 'mega-tumuli' and dome-shaped hills (Ferrarese *et al.* 2002).

References

Ferrarese, F., Macaluso, T., Madonia, G., Plameri, A. and Sauro, U. (2002) Solution and recrystallisation processes and associated landforms in gypsum outcrops of Sicily, *Geomorphology* 49(1), 25–43.

Horta, J.C. de O.S. (1985) Salt heaving in the Sahara, *Géotechnique* 35, 329–337.

A.S. GOUDIE

SALT RELATED LANDFORMS

Evaporites, including halite (sodium chloride) are widespread both geographically and in the geological record. As much as one-quarter of the world's continental areas may be underlain by evaporites of one age or another. When evaporite beds are thick they can have many important geomorphological consequences: the development of diapiric structures (including salt domes), the production of folds and faults (see FAULT AND FAULT SCARP), tectonic uplift and associated drainage modification, the creep of salt as 'salt glaciers', and widespread solution, subsidence and karst formation. Salt is important also as a cause of weathering (see SALT WEATHERING).

Because it has a low density and low rheidity (the ease at which it flows as a viscous solid), salt flows readily under burial conditions when the surrounding sediment is still undeformed. Flow rates can increase in the presence of water (as brine) and at temperatures in excess of around 245 °C. The flowage of salt can transform relatively tabular evaporite bodies into a wide variety of structures that tend to evolve from concordant, low-amplitude features through to discordant high-amplitude intrusions (DIAPIRS), and thence to extrusions. The immature concordant structures include salt anticlines (which have approximately symmetrical cross sections, planar bases and arched roofs), salt rollers (which are also ridge-like, but are asymmetric with a faulted scarp) and salt pillows (periclinal, subsurface domes). Sediment covers tend to be thin over the crests of such structures, but the area above the pillow tends to be a topographic high surrounded by a topographic low or primary rim syncline. These form simultaneously with the accumulation of salt in the area of ongoing uplift, and result from the downwarping of the overlying strata into the space vacated by the salt flowing into the growing structure.

High amplitude diapiric intrusions, involving piercement, with uplift characteristically at 2–4 mm per year, develop in the next stage of structure growth. Among the forms described are salt walls. These are elongated like salt anticlines but are intrusive and of much greater amplitude. They tend to be 4 to 5 km in breadth, have a length of over 120 km and are generally 8 to 10 km apart. Another intrusive form is the salt stock. These vary in shape from squat to columnar and can be conical or barrel-shaped. The round varieties are 2 to 8 km in diametre in their upper parts, and in many places they are linked together in parallel-striking straight or winding lines that have been likened to strings of pearls. When the whole or part of the shallower portion of a diapir extends laterally beyond the cross-sectional area of the diapir roots, an overhang develops, producing balloon or mushroom shapes (see Jackson *et al.* 1990).

The final stage of salt structure evolution is the postdiapir stage (Warren 1989), during which the salt supply is dwindling as the volume of the salt mass decreases. Meteoric processes become important and solutional loss occurs. Large collapse depressions can form.

If the rate of diapiric growth is greater than the rate of salt solution, salt will be extruded at the ground surface. This may tend to be favoured in arid areas where low precipitation values cause low rates of dissolution. Because of its mechanical properties such extruded salt may begin to flow, and such flows are called 'salt glaciers'. These are termed 'namakiers' (from *namak* – the Farsi word for salt – and glacier) (Talbot 1979).

Some of the Zagros Mountains' salt glaciers in Iran are large. The example at Kuh-e-Namak is 2,000 m long, 3,500 m wide and up to 50 m thick, but perhaps the biggest example is Kuh-e-Gach, which is also 50 m thick, but attains a width of around 4,700 m and has an area of around 23.5 km^2. Their speed of movement is less than that of true glaciers, and the average speed is only a few metres per annum with movement tending to occur after rainfall events. They display complex folds where they flow over bedrock irregularities, and at their distal ends they may feather out in a mass of unbedded detritus analogous to a terminal moraine. Salt glaciers are also subject to retreat if the balance of wastage by solution should exceed that of outward flow of salt from the diapir, and this explains the presence of isolated exotic blocks, analogous to erratics, some kilometres beyond the plugs themselves.

The growth of salt structures can create characteristic drainage patterns and slope forms. A model of this has been proposed by Berger and Aghassy (1982) who envisage three development phases. In the 'positive relief stage' there is a central dome on which radial drainage by outbound consequent streams is dominant, and these dissect long isoclinal slopes. In the 'breached stage' the initial topographic high in the centre of the dome becomes lowered, an inversion of topography begins to occur and a major depression develops at the dome's centre. Inward-facing scarp slopes and inbound obsequents develop. In the 'obliterative stage' the inbound obsequents expand headwards and capture much of the consequent outward-bound drainage. Sedimentation, floodplain development and marsh formation occur in the core.

Finally, a large number of the world's 'fold and thrust' belts have developed over evaporites. Examples include the Jura, the Pyrenees, the Franklin Mountains of north-west Canada, the Canadian arctic fold mountains, the Salt Range of Pakistan, the Zagros of Iran, the Sierra Madre Oriental of Mexico, the Cordillera Oriental of Colombia, the Atlas of Tunisia and Algeria, the South Urals, the mountains of the Tadjik Republic, and the anticlinal province of the Amadeus basin in Australia. The presence of salt encourages what is termed 'thin-skinned deformation'.

References

Berger, Z. and Aghassy, J. (1982) Geomorphic manifestation of salt dome stability, in R.G. Craig and J.L. Craft (eds) *Applied Geomorphology*, 72–84, London: Allen and Unwin.

Jackson, M.P.A., Cornelius, R.R., Craig, C.H., Gansser, A., Stöcklin, J. and Talbot, C.J. (1990) Salt diapirs of the Great Kavir, Central Iran, *Geological Society of America Memoir* No. 177.

Talbot, C.J. (1979) Fold trains in a glacier of salt in southern Iran, *Journal of Structural Geology* 1, 5–18.

Warren, J.K. (1989) *Evaporite Sedimentology*, Englewood Cliffs, NJ: Prentice Hall.

A.S. GOUDIE

SALT WEATHERING

The weathering of rocks and building materials by salt. This is an important group of processes particularly in deserts, on coasts and in cities. Salt weathering probably plays an important role in the formation of PANS, TAFONI, STONE PAVEMENTS, SHORE PLATFORMS, rock flour and split cobbles. The build up of salt in rocks can also cause SALT HEAVE OR HALOTURBATION and SLAKING. Conventionally, salt weathering can be divided into mechanical and chemical mechanisms (Goudie and Viles 1997).

The most cited cause of salt weathering is generally the process of salt crystal growth from solutions in rock pores and cracks (Evans 1970). Various mechanisms can cause crystal growth to occur. For example, some salts rapidly decrease in solubility as temperatures fall. This is particularly true of sodium sulphate, sodium carbonate, magnesium sulphate and sodium nitrate. Thus nocturnal cooling could cause salt crystallization to occur. Such a crystallization of a salt solution on a temperature fall affects a much larger volume of salt per unit time than crystallization induced by evaporation, which is a more gradual process.

Nevertheless, evaporation helps to create saturated solutions from which crystallization can occur, and when this happens highly soluble salts will produce large volumes of crystals. In this context it is important to note that of the common salts, gypsum is very much less soluble than many of the others, and that less crystalline material will be available in a given volume of solution to cause rock disruption.

A salt's crystal habit may also affect its power to cause rock breakdown. For instance, the needle-shaped habit of sodium sulphate crystals (mirabilite) might tend to increase their disruptive capability.

Air humidity is an important control of the effectiveness of salt crystallization, for a salt can crystallize only when the ambient relative humidity is lower than the equilibrium relative humidity of the saturated salt solution. If that is the case on a rock surface then the salt will crystallize and cause decay. The equilibrium relative humidities of different salts vary considerably, and those with low values will be prone to dissolution in humid air (Plate 106). The equilibrium relative humidities of hydrated sodium carbonate and sodium sulphate are high, whereas those of sodium chloride, sodium nitrate and calcium chloride are relatively lower.

Another type of salt weathering process is salt hydration. Certain common salts hydrate and dehydrate relatively easily in response to changes in temperature and humidity. As a change of phase takes place to the hydrated form, water is absorbed. This increases the volume of the salt and thus develops pressure against pore walls. Sodium carbonate and sodium sulphate both undergo a volume change in excess of 300 per cent as they hydrate.

For some salts a change of phase may occur at the sorts of temperatures encountered widely in nature; sodium sulphate's transition temperature is 32.4 °C for a pure solution, and falls to 17.9 °C in a NaCl saturated environment. Moreover for some salts the transition may be rapid. At 39 °C the transition from thenardite (Na_2SO_4) to mirabilite ($Na_2SO_4 \cdot 10H_2O$) may take no longer than twenty minutes (Mortensen 1933).

Winkler and Wilhelm (1970) have calculated the hydration pressures of some important common salts at different temperatures and relative humidities, and find that the greatest hydration pressures (maximum value 2,190 atm at 0 °C and 100 per cent relative humidity) occur when anhydrite changes into gypsum. This exceeds the crystallization pressure of ice at −22 °C, and is in excess of the pressure required to exceed the tensile strength of rocks.

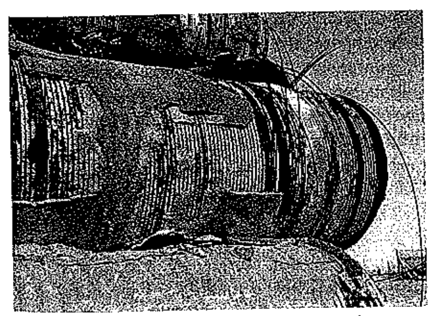


Plate 106 Damaged water pipelines in central Namibia. They proved to be unable to withstand the corrosive conditions associated with the foggy, salty environment of the Namib, and had to be replaced after only a few years in service

The number of occasions upon which rock surface temperatures cycle across the temperature thresholds associated with the change of phase is probably substantial in many desert areas. If one assumes that an air temperature of c.17 °C translates into a rock surface temperature of c.32 °C (the transition temperature for sodium carbonate and sodium sulphate) then that value is crossed daily between 5 and 9 months of the year depending on the desert station selected. In other words, there may typically be around 150 to 270 days in the year in which rock temperature conditions are favourable to the salt hydration mechanism of rock decay.

A third possible mechanism of rock disruption through salt action has been proposed by Cooke and Smalley (1968), who argue that disruption of rock may occur because certain salts have higher coefficients of expansion than do the minerals of the rocks in whose pores they occur. Halite expands by around 0.9 per cent between 0 and 100 °C, whereas the volume expansion of quartz and granites is generally about one-third of that value. Gypsum and sodium nitrate are other common salts that have a relatively greater expansion potential compared to rock minerals.

It is difficult to assess the actual importance of this process, and while some early experimental simulations (e.g. Goudie 1974) suggested that it was not very effective, much more work is required on this mechanism before its potential can be dismissed.

In addition to these three main categories of mechanical effects, salt can cause chemical

Plate 107 Raised beach cobbles being split by halite and nitrate in the Atacama Desert near Iquique in Chile. (Beer can for scale)

weathering. Some saline solutions can have elevated pH levels. Why this is significant is that silica mobility tends to be greatly increased at pH values greater than 9. Indeed, according to various studies, silica solubility increases exponentially above pH9. The presence of sodium chloride may also effect the degree and velocity of quartz solution. At higher NaCl concentrations quartz solubility and the reaction velocity both increase. The growth of salt crystals may be able to cause pressure solution of silicate grains in rocks, for silica solubility increases as pressure is applied to silicate grains. This is a mechanism that has been identified as important in areas where calcite crystals grow, as for example in areas of calcrete formation.

Schiavon *et al.* (1995) have found petrographic evidence from granites in urban atmospheres that suggests chemical reactions occur between granite minerals and weathering solutions responsible for the precipitation of gypsum. They found feldspar minerals that were partially or totally replaced by sulphate crystalline salts while still retaining their primary mineral outline and texture.

The attraction of moisture into the pores of rocks or concrete by hygroscopic salts (e.g. sodium chloride) can accelerate the operation of chemical weathering processes and of frost action (MacInnis and Whiting 1979; Plate 108) and the disruptive action of moisture trapped in rock capillaries is well known.

Plate 108 Concrete buildings in Ras Al Khaimah, United Arab Emirates, illustrating salt weathering caused by migration of salt solutions up the pillars by the wick effect. The process is exacerbated by high ground water levels associated with the spread of irrigation

Salt can have a deleterious impact on iron and concrete. Many engineering structures are made of concrete containing iron reinforcements. The formation of the corrosion products of iron (i.e. rust) causes a volume expansion to occur. If one assumes that the prime composition of such corrosion products is $Fe(OH)_3$, then the volume increase over the uncorroded iron can be fourfold. Thus when rust is formed on the iron reinforcements, pressure is exerted on the surrounding concrete. This may cause the concrete cover over the reinforcements to crack, which in turn permits the ingress of oxygen and moisture which then aggravates the corrosion process. In due course, spalling of concrete takes place, the reinforcements become progressively weaker, and the whole structure may suffer deterioration.

Rates of corrosion are accelerated by chloride ions which may occur in a concrete because of the use of contaminated aggregates or because of penetration from a saline environment. However, the electro-chemical corrosion of metals can also be produced by sulphates for there are often sulphate-reducing bacteria in a saline soil containing sulphates, which can cause strong corrosion to metals.

Sulphates can cause severe damage to, and even complete deterioration of, Portland cement concrete. Although there is controversy as to the exact mechanism of sulphate attack (Cabrera and Plowman 1988), it is generally accepted that the sulphates react with the alumina-bearing phases of the hydrated cement to give a high sulphate form of calcium aluminate known as ettringite. Magnesium sulphate is particularly aggressive because in addition to reacting with the aluminate and calcium hydroxide as do the other sulphates, it decomposes the hydrated calcium silicates and, by continued action, also decomposes calcium sulphoaluminate. The formation of ettringite involves an increase in the volume of the reacting solids, a pressure build up, expansion and, in the most severe cases, cracking and deterioration. The volume change on formation of ettingite is very large, and is even greater than that produced by the hydration of sodium sulphate.

Another mineral formed by sulphates coming into contact with cement is Thaumasite. This causes both expansion and softening of cement and has been seen as a cause of disintegration of rendered brick work and of concrete lining in tunnels. Building materials may also be damaged by SULPHATION, which creates disruptive gypsum on rock surfaces.

Finally, one controversial issue in weathering studies is whether the presence of salts accelerates the rate at which frost action operates, and, if it does, why this should be. Some laboratory studies (see Williams and Robinson 1991) have indicated that some rocks disintegrate more rapidly when they are frozen after soaking in salt solution rather than in pure water, and various studies have shown that de-icing salts can promote the breakdown of concrete by freeze–thaw. However, in some laboratory simulations salts may reduce or even prevent frost weathering.

This whole issue is important in terms of understanding weathering in high latitude coastal situations (e.g. on shore platforms), to understand the effects of de-icing salts on road surfaces and engineering structures like bridges, and also because salts generated by acid rain could conceivably enhance frost action. Williams and Robinson (1991), in a useful review, have looked at some of the mechanisms that might explain why under some conditions salts could accelerate frost weathering.

Identification of areas where salt weathering is a hazard to engineering structures is an important task for applied geomorphologists and a detailed discussion of this in the context of ground water conditions in arid regions is provided by Cooke *et al.* (1982). As irrigation spreads, leading to ground water rise, areas subject to salt attack may increase. There are already many examples of the acceleration of salt weathering by human activities causing the decay of cultural treasures, such as the Sphinx in Cairo, Petra in Jordan and Monhenjo-Daro in Pakistan.

References

Cabrera, J.G. and Plowman, C. (1988) The mechanism and rate of attack of sodium sulphate on cement and cement/pfa pastes, *Advances in Cement Research* 1(3), 171–179.

Cooke, R.U. and Smalley, I.J. (1968) Salt weathering in deserts, *Nature* 220, 1,226–1,227.

Cooke, R.U., Brunsden, D., Doornkamp, J.C. and Jones, D.K.C. (1982) *Urban Geomorphology in Drylands*, Oxford: Oxford University Press.

Evans, I.S. (1970) Salt crystallisation and rock weathering: a review, *Revue de Géomorphologie Dynamique* 19, 153–177.

Goudie, A.S. (1974) Further experimental investigation of rock weathering by salt and other mechanical processes. *Zeitschrift für Geomorphologie Supplementband* 21, 1–12.

Goudie, A.S. and Viles, H.A. (1997) *Salt Weathering Hazards*, Chichester: Wiley.

MacInnis, C. and Whiting, J.D. (1979) The frost resistance of concrete subjected to a deicing agent, *Cement and Concrete Research* 9, 325–336.

Mortensen, H. (1933) Die Salzprengung und ihre Bedeutung für die regional klimatische Gliederung der Wüsten, *Petermanns Geographische Mitteilungen* 79, 130–135.

Schiavon, N., Chiavari, G., Schiavon, G. and Fabbri, D. (1995) Nature and decay effects of urban salting on granite building stones, *Science of the Total Environment* 167, 87–101.

Williams, R.B.G. and Robinson, D.A. (1991) Frost weathering of rocks in the presence of salts – a review, *Permafrost and Periglacial Processes* 2, 347–353.

Winkler, E.M. and Wilhelm, E.J. (1970) Saltburst by hydration pressures in architectural stone in urban atmosphere, *Geological Society of America Bulletin* 81, 567–572.

A.S. GOUDIE

SALTATION

Derived from the Latin *saltare*, saltation, coined by Gilbert (1914), refers to the hopping, jumping or leaping of sediment grains transported by a fluid, whether wind or water (Bagnold 1956). The grains are launched from their bed in a high angle trajectory by lift forces. In air this trajectory is between 20° and 40° downwind. Grains then accelerate in the flow direction as a result of fluid drag. Then, as a result of gravitational and drag forces they fall back to the bed on a more gentle trajectory (10°–15° in air). The entrainment of a grain in water is due to direct fluid lift and drag forces. In air, entrainment can also result from grain collision (Plate 109). The force at which a

Plate 109 Saltating sand grains leaping to a metre or so above the surface during a sand storm at Budha Pushkar in the Rajasthan Desert, India

grain is set in motion (the entrainment threshold), as well as the height and length that it attains in its subsequent jump, is proportional to the shear velocity of the fluid and to grain size. In air, saltation hop-lengths are about 12 to 15 times the height of bounce. Because a sand grain is only two to three times denser than water, the inertia of the rising grain only carries it to a height of about three grain diameters.

In air, on the other hand, saltating grains rise higher, sometimes to several metres, especially after bouncing impacts on rock or pebble surfaces. It is the mode of travel of most wind-blown sand. In addition, the impact of saltating grains can cause a slow downwind movement of grains by surface creep. There is a transition between surface creep, where grains do not lose contact with the bed, and saltation. Some grains may make very short trajectory paths where they barely leave the bed. This process is termed reptation (Rice *et al.* 1995).

Saltation also affects snow and contributes to the development of avalanches (Sato *et al.* 2001).

References

Bagnold, R.A. (1956) The flow of cohesionless grains in fluids, *Philosophical Transactions of the Royal Society of London* A249, 235–297.

Gilbert, G.K. (1914) Transportation of debris by running water, *United States Geological Survey Professional Paper* 85.

Rice, M., Willetts, B. and McEwan, I. (1995) An experimental study of multiple grain-size ejecta by collisions of saltating grains with a flat bed, *Sedimentology* 42, 595–706.

Sato, T., Kosugi, T. and Sato, A. (2001) Saltation-layer structure of drifting snow in wind tunnel, *Annals of Glaciology* 32, 203–208.

A.S. GOUDIE

SALTMARSH

Coastal and estuarine saltmarshes are depositional landforms situated within the upper part of the intertidal zone. Implicit in their definition is the presence of halophytic (salt-tolerant) vegetation. This distinguishes saltmarshes from tidal flats, from which they commonly develop. Inland areas characterized by alkaline soils may also develop a similar vegetation cover. Associated with aridity, these areas are more commonly referred to as salt flat, salt steppe or salt desert and are not considered further in this entry.

Saltmarshes have a wide geographic distribution along temperate and high latitude coasts, but are replaced by MANGROVES SWAMPS in the tropics. Locally, their occurrence is restricted to low wave energy environments which favour the accumulation of fine, generally muddy, sediments. The morphology of most saltmarshes is characterized by a seaward sloping vegetated platform, dissected by networks of tidal channels (also termed 'creeks' or 'sloughs'). The subtle topography of the marsh surface is often associated with a zonation in plant productivity and/or species composition, which results from complex linkages between factors such as salinity stress, nutrient availability, frequency of flooding (a function of elevation) and plant competition.

At a global scale, major differences in marsh character result from the interaction between ecological, climatic, edaphic and hydrographic influences. These are mediated at a regional scale by the nature and abundance of fine sediments and by the range of depositional settings afforded by particular coastal configurations. Saltmarshes are characterized by a particularly strong interplay of physical, biological and geochemical processes and, accordingly, have long been the subject of considerable scientific interest.

Early scientific studies were concerned mainly with the processes of halophyte colonization under the influence of various environmental factors (notably elevation, as the crucial factor determining the frequency of tidal inundation, salinity and soil aeration) and the importance of vegetation (especially dense swards of tall marsh grasses) in the trapping and binding of fine sediment. Research in Europe and North America emphasized the role of coastal halophytes as 'land-building agents', leading to a model of saltmarsh morphological development under the influence of an autogenic plant succession (Chapman 1974). Geographical variations in the plant succession provide one basis for the classification of marsh types (e.g. Adam 1990). Subsequent ecosystem studies have shown that saltmarshes are sites of high biological productivity, the cycling of which is governed by complex vegetation–substrate–fauna interactions and by tidal exchanges of water, sediments and nutrients with the marine environment. The so-called 'outwelling hypothesis' (Odum 1971) attributed much of the enhanced biological productivity of coastal waters to exports of organic material and nutrients from intertidal marshes and subtidal

seagrass beds. Empirical studies provide only partial support for such nutrient exports, and highlight the importance of geomorphological controls on marsh configuration and processes, operating over a range of scales (Nixon 1980).

Geomorphologists have tended to assign a secondary, more opportunistic, role to the colonization of intertidal surfaces by halophytic vegetation. From this perspective, marsh ecological characteristics are viewed as contingent upon the provision of viable substrates for plant colonization. Four main sets of physical factors are implicated: sediment supply; tidal regime; wave climate; and relative sea-level movement (Allen and Pye 1992).

The configuration and extent of coastal margins exert a first-order control on both sediment supply and the 'accommodation space' for saltmarsh development. Frey and Basan (1985), for example, draw attention to physiographic (see PHYSIOGRAPHY) contrasts between the Pacific and Atlantic coasts of North America. On the tectonically active and sediment-deficient Pacific coast, saltmarshes are fragmented and are restricted to narrow fringes around protected embayments, estuaries (see ESTUARY) and (in the north) FJORDS. The more extensive Atlantic coastal plain marshes are continuous over larger areas. Regional variation in the width of the CONTINENTAL SHELF also exerts a control on tidal range which, in turn, defines a zone within which saltmarsh sedimentation can occur.

Saltmarshes are important sinks for fine sediment and play an important role in sediment exchange between estuarine and coastal waters. The nature of saltmarsh sediments varies markedly between *allochthonous* systems, characterized by the deposition of externally derived inorganic sediments, and *autochthonous* systems, dominated by the accumulation of internally produced organic material (Dijkema 1987). The relative importance of inorganic and organic matter accumulation determines the nature of saltmarsh morphodynamic development as well as the ability of both physical and ecological components of the system to adjust to changes in environmental boundary conditions (notably tidal range and mean sea level; French 1994).

Tidal hydrodynamic processes are especially important in controlling the rate and pattern of sedimentation within allochthonous marshes, some of the best-developed examples of which occur under large tidal ranges (e.g. Davidson-Arnott *et al.* 2002). In these systems, the introduction of muddy sediment is controlled by both elevation (which determines the frequency and duration of flooding, or 'hydroperiod') and by proximity to the tidal channels through which most of the tidal water movement occurs (French and Spencer 1993). The feedback between elevation, tidal flooding and sedimentation is a key determinant of long-term marsh morphodynamics (Allen 2000; Friedrichs and Perry 2001). Thus, newly formed marshes typically exhibit rapid rates of vertical sedimentation, whilst the sedimentation is much slower in older marshes, which are higher in elevation and therefore less frequently inundated. Non-tidal inundation, such as that resulting from occasional storm surge events, is of greater importance in introducing sediment to marshes with a very small tidal range.

Wave climate exerts an important local control on horizontal marsh extent, even in otherwise sheltered embayments and estuaries, where small geographical variations in fetch may give rise to significant differences in the character and energetics of the intertidal zone. Wave-induced stresses determine the viability of vegetation establishment, although the influence of waves on the stability of the underlying substrate seems to be of more importance than the mechanical strength of the plants. Wave climate also determines the morphology of the saltmarsh to tidal flat transition. Under moderate wave energies this may take the form of an erosional 'micro-cliff'.

The formation and development of saltmarshes is also related to sea-level change. SEA LEVEL provides a moving boundary condition which, along with tidal range, determines the vertical extent of saltmarsh growth. Modern saltmarshes formed in response to Holocene sea-level rise, and minor oscillations in sea level appear to have been associated with distinct episodes of saltmarsh expansion in areas conducive to fine sediment accumulation. This 'depositional paradigm' (Stevenson *et al.* 1986) has been challenged by the discovery of sedimentary deficits within subsiding deltaic marshes. This has focused attention on the extent to which saltmarshes are able to accumulate sufficient material to keep pace with the forecast rates of sea-level rise under global warming scenarios. In general, contemporary rates of sedimentation within non-deltaic marshes in both North America and Europe exceed present rates of sea-level rise. Furthermore, marsh elevations can adjust to higher rates of sea-level rise through the increased sedimentation which accompanies more frequent inundation (French 1994). This adjustment does, of course, depend on sediment supply. The effect of increased water levels on vegetation and soils is also important, especially in autochthonous marshes with limited inputs of inorganic sediment.

Saltmarshes in many parts of the world have experienced high rates of historical loss through reclamation and destructive industrial uses (such as salt production using evaporation ponds). In addition, large areas of estuarine and open coastal marsh have been lost through recent erosion. This erosion is widely attributed to a combination of contemporary sea-level rise and the presence of seawalls and other structures which restrict any natural landward migration of the intertidal zone. From an ecological perspective, remaining saltmarshes are not only important for the maintenance of estuarine and coastal food chains, but also provide valuable wetland habitats. They also act as a naturally dissipative landform that forms an important element of sustainable coastal defence and flood protection strategies. In particular, a number of studies have shown saltmarsh to be much more effective than unvegetated tidal flats in dissipating wave energy. This function translates into significant engineering cost savings when sea defences are constructed behind a strip of saltmarsh.

These ecological and engineering functions have stimulated efforts to restore saltmarsh, for example through the re-establishment of tidal conditions in reclaimed areas no longer required for agriculture. The success of such schemes has been mixed, and it is now clear that successful engineering of ecological and flood defence functions is crucially dependent upon a sound understanding of the geomorphological processes which act to shape the corresponding natural systems (French and Reed 2001).

References

Adam, P. (1990) *Saltmarsh Ecology*, Cambridge: Cambridge University Press.

Allen, J.R.L. (2000) Morphodynamics of Holocene saltmarshes: a review sketch from the Atlantic and Southern North Sea coasts of Europe, *Quaternary Science Reviews* 19, 1,155–1,231.

Allen, J.R.L. and Pye, K. (1992) Coastal saltmarshes: their nature and importance, in J.R.L. Allen and K. Pye (eds) *Saltmarshes: Morphodynamics, Conservation and Engineering Significance*, 1–18, Cambridge: Cambridge University Press.

Chapman, V.J. (1974) *Salt Marshes and Salt Deserts of the World*, 2nd edition, Lehere: Cramer.

Davidson-Arnott, R.G.D., van Proosdij, D., Ollerhead, J. and Schostak, L. (2002) Hydrodynamics and sedimentation in salt marshes: examples from a macrotidal marsh, Bay of Fundy, *Geomorphology* 48, 208–231.

Dijkema, K.S. (1987) The geography of salt marshes in Europe, *Zeitschrift für Geomorphologie* 31, 489–499.

French, J.R. (1994) Tide-dominated coastal wetlands and accelerated sea-level rise: a NW European perspective, *Journal of Coastal Research Special Issue* 12, 91–101.

French, J.R. and Reed, D.J. (2001) Physical contexts for saltmarsh conservation, In: A. Warren and J.R. French (eds) *Habitat conservation: managing the physical environment*, 179–228, Chichester: Wiley.

French, J.R. and Spencer, T. (1993) Dynamics of sedimentation in a tide-dominated backbarrier saltmarsh, Norfolk, UK, *Marine Geology* 110, 315–331.

Friedrichs, C.T. and Perry, J.E. (2001) Tidal salt-marsh morphodynamics: a synthesis, *Journal of Coastal Research Special Issue* 27, 7–37.

Frey, R.W. and Basan, P.B. (1985) Coastal salt marshes, in R.A. Davis (ed.) *Coastal Sedimentary Environments*, 2nd edition, 225–301 New York: Springer-Verlag.

Odum, E.P. (1971) *Fundamentals of Ecology*, 3rd edition, Philadelphia: Saunders.

Nixon, S.W. (1980) Between coastal marshes and coastal waters – a review of twenty years of speculation and research in the role of salt marshes in estuarine productivity and water chemistry, in R. Hamilton and K.B. McDonald (eds) *Estuarine and Wetland Processes*, 437–525, New York: Plenum.

Stevenson, J.C., Ward, L.G. and Kearney, M.S. (1986) Vertical accretion in marshes with varying rates of sea level rise, in D.A. Wolfe (ed.) *Estuarine Variability*, 241–260, Orlando: Academic Press.

Further reading

Allen, J.R.L. and Pye, K. (eds) (1992) *Saltmarshes: Morphodynamics, Conservation and Engineering Significance*, Cambridge: Cambridge University Press.

SEE ALSO: mangrove swamp; mud flat and muddy coast; tidal creek; tidal delta

J.R. FRENCH

SAND-BED RIVER

An alluvial river in which the bed material is predominantly sand-sized (0.0625–2 mm). Sediment size is a primary control on river form and process and sand-bed rivers therefore exhibit a number of characteristic process and morphological attributes that distinguish them from channels dominated by other sediment sizes. They are recognized by geomorphologists as a fundamental river type, distinct from GRAVEL-BED RIVERS, in which the bed material is coarser (>2 mm) and less well sorted.

Within the drainage network DOWNSTREAM FINING ensures that sand-sized sediments dominate the sediment load delivered to distal reaches. Gravel-bed rivers in upland and piedmont settings therefore give way to sand-bed channels in the lowlands, though sandy reaches will develop wherever sand is supplied in abundance, for example as a function of local lithology or land use.

The grain characteristics and relative importance of SUSPENDED LOAD and BEDLOAD transport vary in sand-bed rivers with sediment supply and flow characteristics, but suspended grains are typically finer than 0.2 mm and account for a greater proportion of total sediment yield. In the limit case, sand-bed rivers can be thought of as suspension-dominated (e.g. Parker in press). Examination of such channels reveals that the boundary shear stresses generated by modest flows (for example bankfull) are at least one order of magnitude larger than the stresses needed to entrain median sizes. This indicates that in sand-bed rivers mass sediment entrainment and transport occur at discharges well below discharges associated with rare floods, and that modest flows are responsible for maximum cumulative sediment yield. Transport conditions are fundamentally different in bedload-dominated, gravel-bed rivers where flows reach entrainment thresholds relatively infrequently. Sand-bed channels are therefore characterized by excess rather than threshold hydraulic stresses, have more mobile and responsive 'live' beds, and carry sediment loads that are limited by supply issues rather than the flow's competence to move available particle sizes.

When stresses are sufficient to entrain grains but turbulent eddying is insufficient to suspend them, grains move as bedload, primarily by SALTATION. Near their entrainment thresholds grains move randomly over a plane bed, but as flow intensity increases patterns of erosion and deposition become ordered in space and groups of sand grains move together as migrating BEDFORMS. Ripples and then dunes form, but a point is reached where structured transport collapses and dunes are replaced by an upper-regime plane bed. With further increases in flow the water surface may develop waves, beneath which antidunes grow. This bedform sequence correlates with both bedload and suspended sediment transport rates and increases of an order of magnitude have been observed between successive stages in flume experiments. The complex mutual adjustments between bedform generation, macroturbulence (burst cycles), flow resistance, and suspended sediment concentrations that ultimately determine hydraulic characteristics and sediment transport rates are incompletely understood, but sufficient has been learned to allow palaeohydraulic interpretation of bedform traces preserved in the modern and lithified deposits of sand-bed rivers.

Because bedform dimensions far exceed grain dimensions in sand-bed channels, a large proportion of total boundary flow resistance is due to bedforms and grain resistance is relatively unimportant. Drag increases as flat beds (grain roughness only) develop ripples then dunes. However, the transition from dunes back to a plane bed is accompanied by a significant drop in resistance that marks a shift from the so-called lower to upper flow regime and ensures that stage-velocity relations are non-linear. Energy losses rise again if the flow becomes supercritical and the water surface develops breaking waves. An additional control on flow velocity in sand-bed channels is suspended sediment concentration, which reduces resistance by dampening turbulence intensity.

The common transition from gravel to sand bed is often abrupt and accompanied by a distinct break of slope. This has been explained in terms of the gradient required to transport sand and gravel loads and the respective importance of channel capacity and competence in suspension and bedload-dominated systems. Slopes in sand-bed reaches are relatively small, typically between 0.002 and 0.0001 (2–0.1 m per kilometre) and exhibit less downstream adjustment than in gravel-bed reaches where large changes in grain size require consequent changes in flow competence.

Cross-section morphology may be regarded as the more adjustable morphological dimension in sand-bed rivers. An element of distinctiveness is apparent in HYDRAULIC GEOMETRY relations for sand-bed channels, but the multivariate nature of the problem makes widespread generalizations difficult. For example, many sand-bed channels transport significant quantities of fine-grained, cohesive sediment (silts and clays) that are deposited in over-bank positions during floods. This facilitates floodplain accretion and increases bank strength, producing channels that have lower width–depth ratios than gravel-bed rivers carrying similar discharges. However, sand-bed channels that do not carry significant quantities of fines have particularly weak banks and tend to be comparatively wide.

Straight, meandering, braided and anabranching rivers may all have sand or gravel beds. Although the patterns are common to both, Lewin and Brewer (2001) suggest that braid and meander formation are fundamentally different in gravel- and sand-bed channels as a function of differences in excess shear stress and thence sediment mobility and bedform persistence. Certainly a degree of distinctiveness is apparent in the controls on planform type. For example, numerous studies have identified threshold values of discharge and channel slope that discriminate meandering and braided forms, but detailed analysis indicates that the threshold slope for braiding also depends on bed material size, with sand-bed rivers tending to occur on more gentle gradients for a given discharge (Knighton 1998: 211). Similarly, examination of the controls on meander geometry suggest that channels carrying greater suspended loads and more cohesive materials (by extension, suspension-dominated sand-bed channels) tend to be more sinuous and have smaller wavelengths than meandering channels in bedload-dominated systems.

References

Knighton, D. (1998) *Fluvial Forms and Processes*, London: Arnold.

Lewin, J. and Brewer, P.A. (2001) Predicting channel patterns, *Geomorphology* 40, 329–339.

Parker, G. (in press) Transport of gravel and sediment mixtures, in *Sedimentation Engineering*, American Society of Civil Engineers, Manual 54.

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.

SEE ALSO: bedform; downstream fining; gravel-bed river; hydraulic geometry; suspended load

STEPHEN RICE

SAND RAMP

Topographically controlled aeolian deposits amalgamated with layers of fluvial, colluvial and talus sediments derived from local mountain sources, and palaeosols representing relatively stable geomorphological periods. Mountains act as barriers to sand transport and sand accumulates on piedmont slopes in their lee and to their windward sides. Where sections are present, multiple periods of sand accumulation and palaeosol development can be identified and such phases can be dated by techniques such as luminescence dating. They can provide a record of past episodes of aeolian activity and stabilization (see, for example, Allchin *et al.* 1978 on the Thar; and Lancaster and Tchakerian 1996 on the Mojave).

References

Allchin, B., Goudie, A.S. and Hegde, K.T.M. (1978) *The Prehistory and Palaeogeography of the Great Indian Desert*, London: Academic Press.

Lancaster, N. and Tchakerian, V.P. (1996) Geomorphology and sediments of sand ramps in the Mojave Desert, *Geomorphology* 17, 151–165.

A.S. GOUDIE

SAND SEA AND DUNEFIELD

Collections of dunes are best called by easily understood terms: 'dunefields', and, when large, 'sand seas'. Many geomorphologists use the term 'erg' for any large collection of dunes (see DUNE, AEOLIAN), but, apart from the indiscriminate use, there are problems with this term. First, 'erg' is an indigenous term from a small part of the northwestern Sahara, where it was used on the topographic maps of over a century ago for anything from a large dune to a very large sand sea. Second, there are many other indigenous terms for bodies of dunes in other parts of the world (even in the northwestern Sahara). And third, better mapping shows there to be two distinct groups of dunefield.

There is a sharp break in the size distribution of collections of dunes at about 32,000 km^2 and the larger group has a sharp peak in size at about 200,000 km^2 (Wilson 1973) (Figure 139). In other words, large dunefields, termed here 'sand seas', are a distinct population. A very similar size break also distinguishes 'seas' in common English usage from smaller bodies of water. The peaked distribution is a function of tectonics. The gentle folds in the basement rock of the African, Asian and Australian deserts are of this size order (Figure 140). Most dunefields and sand

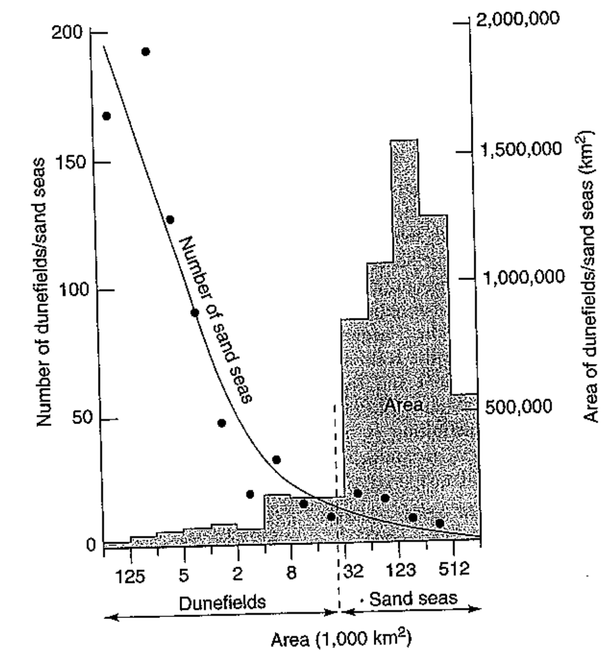


Figure 139 Distribution of the areal extent of aeolian sand bodies, showing the distinction between sand seas and dunefields (after Wilson 1973)

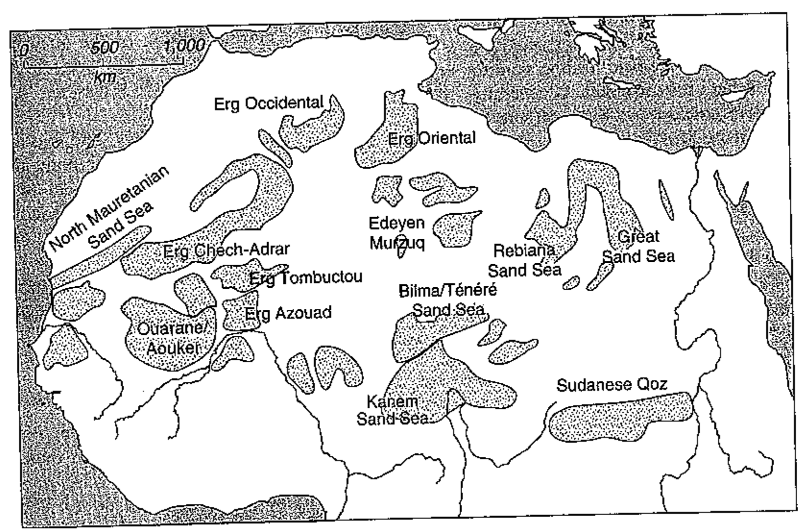


Figure 140 The principal sand seas of the Sahara and Sahel

seas occur in basins or lowlands, and many sand seas nearly fill their basins. Dunefields do so less commonly, although some, as in the Kelso dunefield in California and the Great Sand Dunes in Colorado, fill a large proportion of their small basins.

If the aeolian sand in a dunefield is broken by a patch of desert floor (pavement or bare rock) that is bigger than an interdune, then strictly speaking, this is the edge of the field, but this is a somewhat pedantic restriction. Judgements like this are just one example of the ambiguities in defining the extent of a sand sea (or dunefield), but even with this proviso, the largest active/semi-active sand seas include: the Rub' al Khali in Arabia (about 560,000 km^2) and the Great Eastern Erg in Algeria (about 192,000 km^2). Another somewhat vain definition is of the lower size limit of a dunefield: it could be two dunes.

Sources of sand

Apart from topographic control, the most important control on the distribution of bodies of dunes is proximate supplies of sand. These are of various kinds: weathered bedrock, fluvial deposits, coastal sedimentary cells, and earlier or other dunefields are the most common. Many large dunefields and sand seas derive their sand from a mixture of these sources. In many of the older sand seas, sand has accumulated and been reworked over many cycles. Many dunefields are parts of regional systems of sand movement, one field feeding sand to the next. In the Sahara the movement is generally from north-east to south-west. Regional movements also occur in the Mojave, in the Namib, and probably in Australia. The issue of source, which is best illuminated by studies of mineralogy, has been extensively debated in relation to individual sand seas (Muhs 2003).

Dune assemblages

Many sand seas, unlike most dunefields, contain a variety of dune types. There are several reasons. First, variety is a function of size, for different wind regimes can occur over such large areas, and wind regime is a major determinant of dune type. The northern parts of the Saharan and Arabian sand seas experience frontal winds in winter, and trade winds in summer. Further south in the same sand sea, the influence of the frontal winds fades. In the lower latitude sand seas of southern Africa and Australia, the trade wind systems dominate and dune form is less varied.

Second, variety of dune form is a function of age. Large bodies of sand can only accumulate over many thousands of years, lengths of time that have seen changes in climate. Older dunes may differ in type and orientation from younger ones. Moreover, many older dunes have been subdued by erosion, and have developed soils at times when the climate was wetter. Their gentle topography may be scarred by haphazard reactivation or it may be buried by younger sands to varying degree. This can be seen in the Great Western Erg in Algeria. The northern portion is dominated by ancient, mostly linear, and now subdued dunes, which have developed soils. Further south there are more active and much higher dunes in various formations (transverse, linear and star) (Callot 1988). Another example of how age creates variety is the Wahiba sand sea in Oman. The oldest dunes here have been lithified (see AEOLIANITE), and eroded to an almost level plain. A later generation of large linear dunes partially covers these lithified sands and has moved over them to the north. In the south, these linear dunes have in turn been covered by transverse and network dunes built with new Late Pleistocene sand. These sands have invaded the sand sea from the coast, and are themselves derived from the marine erosion of the ancient lithified aeolianites (Warren 1988). Age, in a sense, allows greater entropy and disorder in dunefields and sand seas.

Third, variety of dune type in large sand seas is a function of the movement of the dune body as a whole. Porter (1986) developed a model for ancient sand seas, now forming aeolian sandstones. It is also to be applicable to some Saharan sand seas. In the model, the finer sand and associated dunes moves forward more quickly than the coarse sand, which remains as a trailing edge of low zibars.

Non-aeolian features

Sand seas and dunefields have some distinctive non-aeolian features. The most striking of these is associated with the blockage of the pre-existing and marginal drainage by the dunes. The drainage and dune water tables feed or have fed many thousands of lakes, some large. These fill up even after the rare storms of today, but were more often full in the wetter periods of the past. These ancient lakes have left deposits ranging from chara limestones to diatomites, sometimes in very thin (5–10 cm thick) drapes on the sand. Round these lakes, in many Saharan and Arabian sand seas, there are signs of human (Palaeolithic and Neolithic) settlement, in the form of tools and middens. Lakes like these are common in the Taklamakan in China and in the Nebraska Sand Sea.

Activity/planetary sand seas

Dunefields and sand seas are in a spectrum from fully active to lithified and largely buried. Some small dunefields are almost wholly active, in the sense that all the dunes are in movement, and most of the surface bare and blowable. But even in the hyper-arid central Sahara and Arabia, many sand seas have portions that are partly stabilized, being remnants of earlier phases bearing remnants of soils. As one moves to the margins of these deserts, as in northern Arabia, or either edge of the Saharan or Australian sand seas, dunes become progressively more stabilized by a cover of vegetation and a developed soil. The climatic limits of aeolian activity are debatable, if only for the reason that a migrating climatic boundary has left a complex legacy of stabilized, reactivated and new dunes. There is no dispute, however, that there are dunefields and sand seas that are now almost wholly inactive, some indeed covered by deep forest. Some of the largest surface sand seas, for example of the proto-Kalahari (2,500,000 km^2), or the Nebraska Sand Hills (57,000 km^2), and the Sudanese *qoz* (about 240,000 km^2) are now almost wholly stabilized. Stabilized dunefields and sand seas also occur in areas in which there are now no deserts, as in northern Canada, the North European Plain, Hungary and northern Tasmania.

Sand seas and dunefields slowly lithify under various influences, and become aeolian sandstones. The sediments of ancient 'ergs' occur from as far back as the Precambrian. The deposits of these sand seas generally show great complexity, as a result of changes in climate and tectonics (Blakey 1988). There are also sand seas and dunefields on Mars, some very similar to terrestrial features (Lancaster and Greeley 1990).

References

Blakey, R.C. (1988) Basin tectonics and erg response, *Sedimentary Geology* 56, 127–151.

Callot, Y. (1988) Evolution polyphasée d'un massif dunaire subtropical: Le Grand Erg occidental (Algérie), *Bulletin de la Société géologique de France*, Série 8, 4, 1,073–1,079.

Lancaster, N. and Greeley, R. (1990) Sediment volume in the North Polar Sand Seas of Mars, *Journal of Geophysical Research – Solid Earth and Planets* 95, 921–927.

Muhs, D.R. (2003) Mineralogical maturity in dune fields of North America, Africa and Australia, *Geomorphology* in press.

Porter, M.L. (1986) Sedimentary record of erg migrations, *Geology* 14, 497–500.

Warren, A. (1988) The dunes of the Wahiba Sands, *Journal of Oman Studies, Special Report* 3, 131–160.

Wilson, I.G. (1973) Ergs, *Sedimentary Geology* 10, 77–106.

Further reading

Cooke, R.U., Warren, A. and Goudie, A.S. (1993) *Desert Geomorphology*, London: University College Press.

Livingstone, I. and Warren, A. (1996) *Aeolian Geomorphology: An Introduction*, Harlow: Addison-Wesley Longman.

SEE ALSO: dune, aeolian

ANDREW WARREN

SANDSHEET

An area of predominantly aeolian sand where dunes with slipfaces are generally absent. Sandsheet surfaces can be rippled or unrippled, and range from flat to regularly undulatory to irregular (Kocurek and Nielson 1986). They form in ergs (sand seas) where conditions are not suitable for dunes, or particular factors act to interfere with dune formation. These factors include a high water table, periodic flooding, surface binding or cementation, the presence of vegetation, and a significant coarse grain-size component. These same factors are also effective in promoting sandsheet accumulation where otherwise only sand transport without deposition would occur.

The classic sandsheet is the Selima Sand Sheet of the Libyan Desert. It covers around 120,000 km^2 and is a largely featureless surface of lag gravels and fine sand broken only by widely separated dunefields and giant ripples. Maxwell and Haynes

(2001) have stressed the role of both fluvial deposition and aeolian modification in its development.

References

Kocurek, G. and Nielson, J. (1986) Conditions favourable for the formation of warm-climate aeolian sandsheets, *Sedimentology* 33, 795–816.

Maxwell, T.A. and Haynes, C.V. Jr (2001) Sandsheet dynamics and Quaternary landscape evolution of the Selima Sand Sheet, Southern Egypt, *Quaternary Science Reviews* 20, 1,623–1,647.

A.S. GOUDIE

SANDSTONE GEOMORPHOLOGY

Landforms developed on sandstone can be arranged in a hierarchical series increasing from the microscopic to the regional scale. Examples of features at these various levels are: (1) etch pits on quartz grains, silica skins; (2) TAFONI, tesselated surfaces, solution runnels; (3) cliffs, domes, towers, arches (see ARCH, NATURAL); (4) CUESTA and scarp assemblages; plateau and canyon assemblages; PSEUDOKARST assemblages, ruiniform assemblages. Explaining features at any of these scales requires an understanding of the variable properties of sandstones.

Sandstones can be most simply defined as clastic rocks in which sand-sized fragments are dominant, but there is considerable variation amongst them, and this variation is of geomorphological significance. The size of the dominant clasts is important, for a fine, silty sandstone responds to erosion differently to a coarse, pebbly one. The proportion of intergranular matrix also is significant, and 'clean' sands or arenites (<15 per cent matrix) need to be distinguished from 'dirty' sands or wackes (>15 per cent matrix). The composition of the grains can vary greatly, for arenites can be divided into quartz (<5 per cent feldspar or rock fragments), lithic (>25 per cent rock fragments), arkose (>25 per cent feldspar) and volcanic (>50 per cent volcanic fragments) subtypes. The amount and composition of intergranular cement, porosity and permeability, and the pattern of bedding and jointing exert important influences on the mechanical and chemical properties of sandstones.

The strength of sandstone depends greatly on its composition. For example, the uniaxial (or unconfined) compressive strength varies from 200 MPa (1 Megapascal = 145 lb/in^2) in a strongly cemented quartz arenite, 20 MPa in a weak sandstone, to 2 MPa in very weakly consolidated sands. Moreover, the strength of sandstones varies greatly with the type of stress applied. Their shearing strength is generally less than half, and their tensile strength only 5 to 15 per cent, of their compressive strength. Furthermore, the strength of saturated sandstone is from 90 to only 10 per cent of the strength of the same rock when dry; the difference between wet and dry strengths increases mainly with the clay content of sandstone. The clay content, together with the degree of cementation, also largely determines the deformability of sandstones. Those which are highly cemented and have little clay tend to fail by brittle fracture, whereas those which are poorly cemented and have substantial clay contents tend to deform elastically before rupturing.

The strength of sandstone mass depends not only on the strength of the intact pieces, but also on their freedom of movement, which, in turn, depends on the spacing, orientation and shearing strength of the discontinuities. Jointing and bedding provide major pathways along which water penetrates sandstones, and thereby influence both weathering of the rock, and the build up of pore-water pressure that may promote failure in the rock.

Strong sandstones generally form major cliffs, but the modes of failure on cliffs, and therefore the morphology of the cliffs themselves, varies with the characteristics of particular rock masses. Because sandstones are weakest when in tension, horizontal projections of rock from cliffs, unless in the form of supported arches, generally extend no more than 10 to 20 m. The patterns of scars caused by tensional rupturing indicate that about 50 per cent of the failures through intact sandstone are generated upwards, about 40 per cent laterally, and, contrary to the predictions of commonly used models, only about 10 per cent downwards. In well-jointed sandstones, the dominant mechanism of failure is the collapse of individual blocks that have been undercut. Once undercutting has penetrated beneath the central third of a block, the block will generally topple outwards, unless held in place by pressure exerted by adjoining blocks.

Undercutting of sandstone is mainly the result of the breakdown of less resistant, underlying rocks. While seepage promotes the plastic failure of clay-rich rocks such as shales, even in these rocks, failure frequently takes the form of brittle

fracture along closely spaced beds and joints. Undercutting can also occur along prominent bedding planes in the sandstone itself, and in beds of conglomerate that commonly occur at the base of upward fining cycles of sandstone deposition. Seepage promotes undercutting, especially in very permeable sandstone, but undercutting is essentially the result of the very high concentration of stress generated at the base of a cliff by the weight of the rock above. As the rock at the base is compressed, it deforms laterally, and when the lateral stress pushing outwards into the undercut exceeds the tensional strength, the rock fractures. Tensional stresses generated in this fashion at the base of sandstone cliffs can be greatly augmented by active, or residual, tectonic stresses. For example, the measured horizontal stresses south of Sydney, Australia, are three times greater than the vertical stress. In these conditions, zones of tension, in which joints separating the blocks are opened, extend up the entire cliff face and onto the edge of the adjacent plateau surface. Where sandstones are tilted, joint-bounded blocks may topple or slide down the slope. High and narrow blocks tend to topple; wide and flat blocks tend to slide. Movement of blocks depends also on the steepness of the slope and the frictional resistance at the base of a block.

The form of cliffs thus depends not only on the ROCK MASS STRENGTH of the sandstone itself, but also on the properties of the rocks exposed beneath it. Along the Arnhemland Escarpment, in northern Australia, the Kombolgie Sandstone stands in vertical cliffs where less resistant schists and granites are exposed beneath it, but forms irregular slopes of lower declivity where it occupies the full height of the escarpment. Armouring of weaker rocks by debris from sandstone outcrops also retards undercutting. In the south-west of the USA the presence or absence of thick mantles of talus seems to determine whether slopes are in equilibrium with the properties of the sandstone, or whether they are controlled by foot-slope processes and undercutting.

Plate 110 Cliffs with cavernous weathering in the Nowra Sandstone, south-east Australia

Extensive masses of sandstone may be incorporated in major failures in weaker, underlying rocks (see TOREVA BLOCK). Major rotational failures in clays on the southern part of the Msak Mallat Escarpment, in central Libya, have incorporated slabs of the Nubian Sandstone in a 3-km wide belt of mounds over 100 m high. Removal of confining stresses resulting from the incision of the Cataract Canyon of the Colorado River has allowed evaporites to slowly deform down the dip, causing fracturing and collapse of the overlying sandstones. The result is a series of spectacular graben-like depressions, which are 150 to 200 m wide and are 25 to 75 m deep. Gliding of sandstone blocks away from cliff lines is not limited to areas of highly deformable substrate. South of Sydney, Australia, the movement of large blocks of Nowra Sandstone away from cliff lines is the result of the very slow creep of underlying sandy siltstone (Plate 110).

Many outcrops of sandstone have been shaped into domes and rounded slopes, known as slickrock. In some cases curved sheeting in the sandstone roughly parallels the rounded topography. Most rounded slopes, however, appear to be the result of the granular breakdown of the sandstone and of the peeling of thin, weathered layers from the surface. This is so both in arkosic rocks, such as the Navajo Sandstone of the Colorado Plateau, and in quartz arenites, such as those which are cut into complex arrays of rounded towers in the Bungle Bungle Range of north-west Australia. Granular breakdown and slab failure also are the primary processes in the development of arches, which are quite common in some sandstones. The domes and slickrock slopes that are characteristic of many coarse conglomerates can likewise be attributed to granular breakdown and slab failure.

Weathering of sandstones varies with the mineralogy of its constituents. Where there is considerable matrix, hydration of clay is important. In arkosic sandstone, the primary process is the weathering of feldspars. Even highly quartzose sandstones are subject to extensive, though slow, CHEMICAL WEATHERING, with the order of decay generally being first the siliceous, intergranular cement, then the quartz grains, and finally the quartz overgrowths. The presence of sodium chloride, either as sea spray or as aerosols in rain, has a strong accelerating effect on the dissolution rates of silica. These various processes are dependent on the ability of water to penetrate the sandstone. The opening of joints and generation of microfractures by initial release of confining pressures create numerous pathways for seepage. Even the alteration from fresh to slightly weathered sandstone can result in a reduction by a factor of two or three in mean strength, and of about six in deformability, as porosity and moisture content increase. The advance of weathering depends very much on the permeability of the sandstone, for, where the original pore spaces are filled by overgrowths or siliceous cement, active weathering is essentially limited to a thin surface layer. Nevertheless, prolonged weathering can produce deep solutional features even in quartzites.

The removal of the more soluble constituents in sandstone, and the precipitation of siliceous and iron-rich materials, may result in case hardening of the surface layer. Some case-hardened surfaces develop distinctive polygonal patterns of cracks, known as 'elephant skin'. The cracking seems to be the result of fatigue in the case-hardened layers and of changing surface stresses, similar to the crazing of glazed pottery surfaces. Fire is an important agent in the breakdown of sandstone surfaces in well-vegetated areas. Near Sydney, such surfaces are likely to be exposed to fire once every ten to twenty years. Fire moving rapidly over sandstone produces minor spalling and granular disintegration, while intense fires may cause spalling to depths of 2 cm. Sand blasting of sandstone surfaces during high winds has been overrated, especially as a cause of CAVERNOUS WEATHERING in humid areas, but is very important in arid lands. Shattering by crystal pressures developed in fractures due to freezing is particularly important under cold climates, especially in quartzites, which, though very resistant to abrasion, are brittle rocks that can withstand only slight strain deformation before they rupture explosively.

Initial classifications of sandstone landforms at the regional scale were based on the presumed controlling effect of climate. Climate is undoubtedly the dominant control in the amazing wind eroded YARDANG landscape of the hyper-arid Boukou region of the Sahara. Pseudokarst assemblages, like the impressive array of deep dolines and caves etched into the quartzites of the Roraima region of Venezuela, are certainly best developed in the humid tropics. But solutional features occur extensively on sandstones outside the tropics. Furthermore, some of the most common types of sandstone terrain, such as rounded slickrock relief, and the angular towers and great cliffs of ruiniform assemblages, occur under various climates. The variable properties of the rock, rather than climate, may thus be of primary importance in the shaping of sandstone landforms.

Further reading

Howard, A.D., Kochel, R.C. and Holt, H.E. (eds) (1988) *Sapping Features of the Colorado Plateau, A Comparative Planetary Geology Field Guide*, Washington: NASA.

McNally, G.H. and McQueen, L.B. (2000) The engineering properties of sandstones and what they mean, in G.H. McNally and B.J. Franklin (eds) *Sandstone City, Sydney's Dimension Stone and Other Sandstone Geomaterials*, 178–196, Sydney: Geological Society of Australia.

Mainguet, M. (1972) *Le modelé des grès: Problèmes généraux*, Paris: Institut Geographique National.

Oberlander, T.M. (1977) Origin of segmented cliffs in massive sandstones of southeastern Utah, in D.O. Doehring (ed.) *Geomorphology of Arid Regions*, 79–114, Boston: Allen and Unwin.

Wray, R.A.L. (1997) A global review of solutional weathering forms on quartz sandstone, *Earth Science Review* 42, 137–160.

Young, R.W. and Young, A.R.M. (1992) *Sandstone Landforms*, Berlin: Springer.

R.W. YOUNG

SAPROLITE

Saprolite is weathered rock in which the fabric of the original rock is preserved. The word derives from the Greek sapros (σαπρος or σαπροσ) meaning rotten and was coined in 1895 by Becker. The term is generally applied to chemically altered rocks where all or part of the primary minerals are changed to new-formed minerals. Most often saprolite is referred to granitic rocks, but the term applies to the weathered component of any rock type.

The body of *in situ* weathered rock referred to as saprolite may comprise a number of zones (horizons, layers) depending on the relative rates of weathering, erosion and the composition and

hydromorphic characteristics of the regolith. Figure 141 summarizes data on saprolite and various terms used by various authors for various parts of weathering profiles and their saprolite component.

Starting from the bottom of a weathering profile *saprock* is the part of the profile closest to the weathering front (Figure 141). Saprock is rock that has begun to weather, but only about 20–30 per cent of the primary minerals are chemically altered. This is hard to estimate and a preferable definition is that the material requires a sharp hammer blow to break it. The zone of saprock may contain boulders, beds or other masses of unweathered bedrock. At the weathering front, which may be very sharp or gradual, rocks change from fresh to partly weathered, alteration of feldspars to clay minerals is seen and ferromagnesian minerals release Fe^{2+} that is oxidized to red-yellow coloured Fe^{3+} oxihydroxides or oxides.

The form of the WEATHERING FRONT may be relatively flat or very irregular, the latter being more common. Its shape is primarily dependent on the nature of the rocks being weathered. In massive igneous bodies the form is generally more regular – but with isolated sub-spherical boulders or corestones of fresh rock in the saprock – than that in dipping sedimentary rocks or metamorphic rocks. The number and orientation of joints, cleavage and bedding planes that form the initial conduits for the weathering solutions mainly control this. Plates 111–115 are examples of different types of saprolite and weathering fronts.

Moving up profile the saprock gradually gives way to saprolite where the majority of labile (more easily chemically weathered) minerals are altered and replaced by new minerals formed by the chemical weathering. The most common minerals in saprolite are clays, iron oxihydroxides and oxides and any minerals largely resistant to weathering (e.g. quartz, magnetite or pre-existing clay minerals). The clay minerals generally gradually change from smectite (Si/Al = < 1) lower in the profile, where drainage is generally poorer than higher where the dominant clay is kaolinite (Si/Al = 1). Kaolinite forms, as drainage of the weathering solutions is more efficient.

Above the saprolite is collapsed saprolite or the 'mobile zone'. In this zone (Figure 141) the

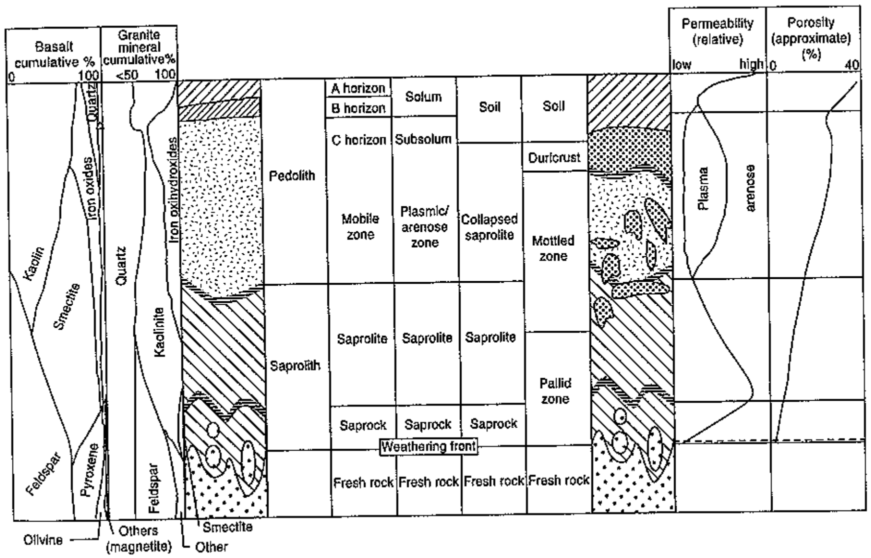


Figure 141 Some of the various terminologies used to describe weathering profiles including saprolite. On the left are two sketch examples of how mineralogy varies through the saprolite and on the right is an example of how the hydraulic properties of the weathering profile and saprolite vary with depth. (modified from Taylor and Eggleton 2001)

Plate 111 A small erosional remnant of a 'typical' weathering profile developed in felsic gneiss on the Yilgarn near Kalgoorlie in Western Australia. It is about 3 m high and shows ferruginous cap about 0.25 m developed over a 0.5 m thick mottled saprolite that grades downward into 3 m of gneissic saprolite. The saprolite is composed of relict quartz grains, and kaolinite with minor haematite. The composition of the mottled saprolite is much the same with increased quantities of haematite. The ferruginous crust is mainly haematite with quartz and minor goethite (photo Ian Roach)

Plate 113 A very deep weathering profile formed in ultra-mafic rocks at Marlborough in central Queensland, Australia. About 40 m of the profile is exposed, but no fresh rock occurs in this pit. The total profile depth is up to about 100 m. The bulk of the material in the photo is saprolite composed of clay minerals, mainly nontronite and talc with Fe^{3+} oxides, goethite and haematite and secondary silica as chrysoprase and chalcedony. The upper part of the profile is transported debris that in places has filled karstic channels in the upper saprolite. The hill is capped by siliceous duricrust derived from silica mobilization during the weathering of the ultra-mafic rocks

Plate 112 A deep weathering profile on the Yilgarn Craton of Western Australia showing similar features to those described in Plate 111. The profile here is some 30 m thick and at the base of the pit an irregular weathering front between fresh and weathered felsic granites can be observed. The original rock structures (joints) are clearly preserved in the saprolite in this profile

Plate 114 A 10 m section through a 20 m weathering profile formed in Ordovician intermediate volcanic rocks at Northparkes in New South Wales, Australia. This profile is unusual in that the pallid or bleached saprolite occurs above the mottled zone in the right of the photograph. It comprises mainly kaolinite in the bleached zone with minor secondary quartz and calcite and gypsum. The lower mottled zone has similar mineralogy with the addition of haematitic mottles. These mottles have been dated as Carboniferous (Pillans *et al.* 1999). The saprolite has been partly eroded and covered in part by palaeochannel sediments (left of photo) that are themselves weathered significantly and now form sedimentary saprolite with large haematitic and goethitic mottles. These mottles date from the Miocene. The whole profile is overlain by up to 2 m of Quaternary alluvium with a red-brown earth formed in it

Plate 115 A 7-m thick section through Proterozoic quartzite unconformably overlain by saprolite formed from Cretaceous labile sandstones and Quaternary transported cover at Darwin, Australia. The Proterozoic rocks are fresh. The Cretaceous has been completely transformed to a quartz/kaolinite saprolite and the whole sequence overlain by ferruginous lag gravel, sand and clays. It is interesting that the upper surface of the saprolite is karstic and that the overlying lags, etc. have filled the karstic channels

saprolite has been chemically eroded (weathered) to such a high degree that the original rock fabric is no longer self-supporting and it has collapsed. This process of collapse may also be enhanced by processes of bioturbation (e.g. tree roots, termites) and/or pedoturbation (e.g. wetting and drying, and illuviation). It is also from this point in the profile that the REGOLITH may move down slope (Plate 116) under the influence of gravity. Also within this zone, quartz sand and clays may begin to separate into clots of clay surrounded by sand (Taylor and Eggleton 2001: 161) by processes of pedoturbation.

Plate 116 Overturned Proterozoic metasediments moving down a slope of about 1.5° at the Mary River, Northern Territory, Australia

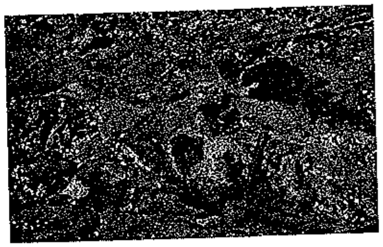


Plate 117 Close-up photograph of mottles on a wave cut platform at Darwin, Australia. The hammer provides scale. These mottles are up to 15 cm across and are composed of haematite, providing the red colour, cementing a saprolitic matrix of kaolinite and quartz. The intervening bleached saprolite is identical, except that it does not contain haematite

Secondary overprints may affect the appearance of the weathering profile and the saprolite. The most widely observed modification is the formation of iron oxihydroxide aggregations called mottles. Mottles (Plate 117) may extend throughout the profile, but are mostly observed in the upper collapsed saprolite. Another common feature in profiles developed mainly on felsic rocks is a bleaching or the removal of Fe-oxihydroxides from the lower parts of the saprolite and saprock. This zone is referred to by many as a pallid zone, particularly in what are described by some geomorphologists as a 'lateritic profile'.

Figure 141 (p. 909) shows idealized examples of how some physical and mineralogical properties of saprolite change through a profile. One point of interest here is the commonly observed addition of quartz to the upper parts of saprolite developed from basalt, which of course contains no quartz when fresh. This indicates the addition of 'foreign' material to saprolite profiles may occur either by overtopping of the saprolite or by other deposits from aeolian accession. This is a very common feature of most saprolite profiles but is most readily observable over weathered mafic rocks.

References

Becker, G.F. (1895) Gold fields of the southern Appalachians, *Annual Report of the United States Geological Survey, Part III. Mineral Resources of the United States, Metallic Products* 16, 251–331.

Pillans, B., Tonui, E. and Idnurm, M. (1999) Palaeomagnetic dating of weathered regolith at Northparkes Mine, NSW, in G. Taylor and C.F. Pain (eds) *Regolith '98, New Approaches to an old Continent, Proceedings*, 237–242, Perth: CRC LEME.

Taylor, G. and Eggleton, R.A. (2001) *Regolith Geology and Geomorphology*, Chichester: Wiley.

GRAHAM TAYLOR

SASTRUGI

Sharp irregular ridges, mounds or dunes. They form on ice sheets, ice caps, sea ice and tundra (typical of Antarctica and Greenland), and are composed of ice and compacted snow. Originating from the Russian word *zastrugi*, they are formed by aeolian erosion and the deposition of drifting snow (Gow 1965), and typically are 1–2 m long and 10–15 cm in height (though exceptional cases can reach 1.5 m in height and hundreds of metres long). Sastrugi align longitudinally with the predominant wind direction, making it possible to infer the prevailing wind direction at the time of sastrugi development from their configuration. They often form after a blizzard on the hard ice surface, becoming larger and harder as the blizzards blow across them throughout the winter months. Sastrugi are usually found in the lee of obstacles but are also known to exist in open conditions.

Reference

Gow, A.J. (1965) On the accumulation and seasonal stratification of snow at the South Pole, *Journal of Glaciology* 5, 467–477.

Further reading

Warren, S.G. and Brandt, R.E. (1998) Effect of surface roughness on bidirectional reflectance of Antarctic snow, *Journal of Geophysical Research – E: Planets* 103 (E11), 25,779–25,788.

STEVE WARD

SCABLAND

A scabland is an erosional landscape formed by a catastrophic flood and is generally applied to the effects of Jökulhlaups. It was first introduced by Bretz (1923) to describe the erosion and stripping of the Columbia Plateau Basalts by floods from Glacial Lake Missoula in eastern Washington, USA. Bretz adopted a term that had been used by local farmers to describe the 'scabby' terrain: 'The terms "scabland" and "scabrock" are used in the Pacific Northwest to describe areas where denudation has removed or prevented the accumulation of a mantle of soil, and the underlying

Table 41 Bedforms in the Channeled Scabland

	Scoured in rock	Scoured in sediment	Depositional
Macroforms (scale controlled by channel width)	Pool and riffle sequence, Quadrilateral residual forms in channel Anastomosis	Large-scale streamlined residual forms	Longitudinal bars (a) Pendant bars (b) Alternate bars (c) Expansion bars Eddy bars
Mesoforms (scale controlled by channel depth)	Longitudinal grooves Pot-holes Inner channels Cataracts	Scour marks	Large-scale transverse ripples (giant current ripples)
Microforms	Scallop pits	Not preserved	Small-scale ripple stratification (restricted to slack water facies)

Source: From Baker (1978a)

rock is exposed or covered largely with its own coarse, angular debris' (Bretz 1923: 617).

Channeled Scabland

The formal physiographical region known as the 'Channeled Scabland' is located in the northern portion of the Columbia Plateau in eastern Washington, USA and comprises an area of approximately 40,000 km^2. It is a spectacular channel complex eroded deeply into loess and basalt bedrock. The large flood discharges spilled over pre-flood divides into adjacent valleys and produced the effect of channels dividing and rejoining to form anastomosing (see ANABRANCHING AND ANASTOMOSING RIVER) complexes. These divide crossings are several hundred feet above valley floors.

A typical scabland complex includes erosional and depositional forms. Baker (1978a) has adopted a hierarchical classification of bedforms for the Channeled Scabland (see Table 41). The erosional landforms include grooves, pot-holes, rock basins, inner channels and cataracts. Bretz *et al.* (1956) ascribed several scabland features to differences between various basalt flows. Cataracts, such as Dry Falls, formed as one group of basalt flows was stripped from underlying resistant flows. Where they were exposed by the floods, the columnar jointed basalts exerted a strong joint control and the plucking action by floodwater yielded boulders >30 m diameter. The most spectacular of the depositional forms are the streamlined channel deposits, some superimposed by giant current ripples 0.5 to 7 m high and 18 to 130 m in chord length. They are composed predominantly of gravel and boulders. Slackwater deposits accumulated in low velocity areas including re-entrants to the major valleys and in pre-flood tributary valleys.

The scablands were formed by discrete outbursts from a range of sources. These included an enormous subglacial reservoir that extended over much of central British Columbia (conservative estimates of water volume are 10^5 km^3) (Shaw *et al.* 1999) and Glacial Lake Missoula. This lake impounded 2,184 km^3 of water during its maximum extent (O'Connor and Baker 1992). The last major period of scabland flooding is placed approximately between 18,000 and 13,000 years BP. Facies analysis of sedimentary sequences suggest that there may have been as many as forty floods (Waitt 1985) each separated by decades or centuries. Shaw *et al.* (1999) propose that there were fewer floods and that many of the variations in the sedimentary sequences can be ascribed to pulses within a flood caused by the input of multiple sources of floodwaters during these long duration flows (up to 1,000 days).

High water marks along the scabland channels have been used to reconstruct the maximum flood stages and water surface gradients. These include eroded channel margins, depositional features, ice rafted ERRATICS and divide crossings. Discharges as large as 21.3×10^6 m^3 sec^{-1} were conveyed through the channel scabland (Baker 1978b). Some constricted channels reached velocities as high as 30 m sec^{-1}. These high velocities were

Plate 118 Portion of MOC image M2101914 which is centred near 7.89°N, 153.95°E, pixel resolution is 4.4 m. (Malin *et al.* 2001). This image shows anastomosing channel pattern and multiple streamlined forms in a flood channel emanating from a fissure. For detailed description see Burr *et al.* (2002)

possible because of the combinat[illegible] of great flow depth (60 to 120 m) and very steep water surface gradients 2 to 12 m/km.

On Mars, data from orbiting satellites have detected stripped zones on the floors of outflow channels in the Chryse Basin. These anastomosing complexes are 100 km wide and flow over 2,000 km across the planet's surface. They are usually initiated from collapsed zones. Other examples show multiple and asynchronous flows emanating from geothermal fissures in recent Martian history (see Plate 118) (Burr *et al.* 2002). By analogy to the scablands on Earth, it is generally accepted that the Martian outflow channels were also formed by catastrophic floods. Martian outflow channels include a distinctive assemblage of scabland landforms: regional and local anastomosing patterns, expanding and contracting reaches associated with flow constrictions, streamlined hills, inner channels with recessional headcuts, pendant forms (bars or erosional residuals) on the down current sides of flow obstacles, longitudinal grooves, irregular 'etched' zones on channel floors and scour marks around obstacles (Baker 1982).

References

Baker, V. R. (1978a) Large-scale erosional and depositional features of the Channeled Scabland, in V.R. Baker and D. Nummedal (eds) *The Channeled Scabland*, 81–115, Washington, DC: NASA.

—— (1978b) Paleohydraulics and hydrodynamics of Scabland floods, in V.R. Baker and D. Nummedal (eds) *The Channeled Scabland*, 59–79, Washington, DC: NASA.

——(1982) *The Channels of Mars*, Austin: University of Arizona Press.

Bretz, J.H. (1923) The channeled scablands of the Columbia Plateau, *Journal of Geology* 31, 617–649.

Bretz, J.H., Smith, H.T.U. and Neff, G.E. (1956) Channeled Scabland of Washington; new data and interpretations, *Geological Society of America Bulletin* 5, 957–1,049.

Burr, D.M., Grier, J.A., McEwen, A.S. and Keszthelyi, P. (2002) Repeated aqueous flooding from the Cereberus Fossae: evidence for very recently extant, deep groundwater on Mars, *Icarus* 159, 53–73.

Malin, M.C., Edgett, K.S., Carr, M.H., Danielson, G.E., Davies, M.E., Hartmann, W.K., *et al.* (2001) *M21-01914*, Malin Space Science Systems Mars Orbiter Camera Image Gallery (http://www.msss.com/moc_gallery/). (http://photojournal.jpl.nasa.gov/).

O'Connor, J.E. and Baker, V.R. (1992) Magnitudes and implications of peak discharges from Glacial Lake Missoula, *Geological Society of America Bulletin* 104, 267–279.

Shaw, J., Munro-Stasiuk, M., Sawyer, B., Beaney, C., Lesemann, J., Musacchio, A., Rains, B. and Young, R.R. (1999) The channeled scabland: back to [illegible]

[illegible]att, R.[illegible] (1985) Case for periodic, colossal jokulhlaups from Pleistocene glacial Lake Missoula, *Geological Society of America Bulletin* 96, 1,271–1,286.

MARY C. BOURKE

SCANNING ELECTRON MICROSCOPY

The Scanning Electron Microscope (SEM), sometimes used in association with Energy Dispersive Spectrometry (EDS), has been used for studying the surface textures (and chemistry, with EDS) of sediments (especially quartz grains) since 1962. The use of the SEM has had many implications for geomorphology, including the determination of the origin of depositional landforms, the provenance of sediments, the energy of environments and processes of diagenesis and weathering and their development through time. Examples of the use of the SEM include the separation of till from glacilacustrine and glacifluvial grains within the glacial sedimentary environment, studies of the origin of fine silt and clay particles in the geological column, examination of aeolian and other environmental fracture–abrasion mechanisms in the field and the laboratory, and analysis of grain modification under different weathering regimes. A full discussion of grain textures associated with aeolian, fluvial, mass wasting, glacial, tectonic, impact, weathering and diagenetic processes is provided by Mahaney (2002).

Reference

Mahaney, W.C. (2002) *Atlas of Sand Grain Surface Textures and Applications*, New York: Oxford University Press.

A.S. GOUDIE

SCHMIDT HAMMER

A concrete test hammer originally designed by E. Schmidt in 1948 for carrying out *in situ* tests on the hardness of concrete. The instrument measures the distance of rebound of a controlled impact on a rock surface. Because elastic recovery (the distance of repulsion of an elastic mass upon impact) depends on the hardness of the surface, and hardness is related to mechanical strength, the distance of rebound (R) gives a relative measure of surface hardness or strength.

In the Schmidt Type N hammer (which weighs 2.3 kg) the energy of impact (0.224 mkg) is obtained by releasing a spring-controlled plunger. The R value is shown by a pointer on a scale on the side of the instrument (range 10–100). This value represents the rebound distance as a percentage of the forward movement.

The Schmidt Hammer Type N is light and portable and allows *in situ* tests to be made in the field. By enabling quantitative comparison of the hardness of materials it provides a useful tool for geomorphologists. Among its uses have been the description of *nari* (calcrete) profiles, case hardening on tropical karst surfaces, and various types of aggregate resources (Day and Goudie 1977). It has also been much used to assess degree of weathering and the ages of geomorphic features upon which weathering phenomena occur (Ballantyne *et al.* 1989; McCarroll 1991).

Schmidt hammer rebound values have been found to correlate well with other measures of rock strength, including Young's Modulus and Uniaxial Strength (Katz *et al.* 2000)

References

Ballantyne, C.K., Black, N.M. and Finlay, D.P. (1989) Enhanced boulder weathering under late-lying snowpatches, *Earth Surface Processes and Landforms* 14, 745–750.

Day, M.J. and Goudie, A.S. (1977) Field assessment of rock hardness using the Schmidt Test Hammer, *BGRG Technical Bulletin* 18, 19–29.

Katz, O., Reches, Z. and Roegiers, J.-C. (2000) Evaluation of mechanical rock properties using a Schmidt Hammer, *International Journal of Rock Mechanics and Mining Sciences* 27, 723–328.

McCarroll, D. (1991) The Schmidt Hammer, weathering and rock surface roughness, *Earth Surface Processes and Landform* 17, 477–480.

A.S. GOUDIE

SCREE

The terms scree and talus are synonymous, the former being preferred in England and the latter (equivalent to the French word for slope) used predominantly in North America and elsewhere. Both terms describe accumulations of loose, coarse, usually angular rock debris at the foot of steep rock slopes. The terms are used to describe both the landform and the material of which it is composed. The debris accumulations forming screes must be of sufficient thickness to develop a characteristic morphology independent of the underlying slope. Simple debris veneers only a few particles thick are termed 'debris mantled slopes' (Church *et al.* 1979).

Scree slopes occur in a wide range of environments but most significantly in environments where physical weathering processes dominate. The production and/or accumulation of debris must be greater than its subsequent weathering or removal. The coarseness of most scree deposits makes them resistant to subsequent erosion: they are often stable long-lasting elements of the landscape and may be preserved as fossil forms.

The basic characteristics of scree slopes depend primarily on the morphology (and thereby geology) of the flanking cliffs and the geomorphic processes involved in their development. Although the dominant process is usually assumed to be ROCKFALL, scree modification and accumulation may be the result of several different processes acting singly or in combination and several distinctive types of scree may be recognized. The plan morphology of screes depends on the form of the cliffs supplying the debris and the morphology of the surface on which the debris accumulates. Relatively simple cliffs, straight in plan view, tend to produce sheet (straight) rockfall talus slopes lacking significant lateral variation in their characteristics. As cliffs become more dissected, deposition is focused below couloirs (gullies) leading to the development of cones. As well as funnelling rockfalls, couloirs channel other rockwall processes (surface stream flow, snow avalanches (see AVALANCHE, SNOW), etc.) down the cliff, sometimes resulting in significant modification of the scree below. Therefore the cones developed below dissected cliffs are rarely simple single-process forms. In alpine environments debris cones can be significantly modified by snow avalanche activity. DEBRIS FLOW generation may also occur during heavy rainstorms when drainage from the cliff zone is focused by gullies onto finer grained materials at the head of the scree.

Rockfall screes result from the accumulation of discrete rockfall events over long periods of time. The basic characteristics are fall sorting (a logarithmic increase in mean grain size downslope) and a straight slope, often with a well-developed basal concavity. There is continuing debate about whether these straight slopes reflect the angle of repose (see RESPOSE, ANGLE OF) of coarse cohesionless material at about 35° (Carson 1977). Measured profiles of the upper part of many rockfall screes range between 32° and 37° but locally reach 40°. The degree of basal concavity varies with the length of slope and the characteristics of the basal zone.

Fall sorting on rockfall scree slopes, though not universal, primarily results from two mechanisms. Larger boulders have greater momentum and therefore tend to travel further downslope. Second, the frictional resistance (roughness) of the surface over which the boulder slides, rolls or bounces is defined by the relationship between the size of the moving boulder and the irregularities of the surface (boulders and voids) over which it is passing. Big boulders are only effectively trapped in large 'holes' or when they impact other large boulders. The degree of sorting depends on the slope length, cliff height and the size and shape of dominant particles. Locally random effects or differences in particle shapes may result in the absence or anomalous patterns of sorting. On scree slopes modified by snow avalanches loose surface material is swept from the upper slopes to the end of the avalanche track at or beyond the base of the scree, in extreme cases forming AVALANCHE BOULDER TONGUES. These screes show little size sorting on upper slopes but a rapid increase in mean grain size towards the base despite the presence of an unstable scattering of loose rock debris on the surface of larger clasts. Avalanche modified screes have strong basal concavities. Many scree cones have gullies formed by fluvial activity at their head and may have significant debris flow forms (levees, channels and terminal lobes) extending across the scree slope. In extreme cases these have been termed 'alluvial talus'. Scree-like forms produced by the breakup of single large rockfalls generally lack fall sorting and have more complex long profiles. In alpine areas large multiprocess scree cones may display complex surface characteristics associated with the interaction of these processes (Plate 119).

Most scree slopes, despite their coarse debris veneer, have considerable quantities of interstitial fine material at depth or exposed on the higher parts of the slope. They are not exclusively formed of coarse cohesionless material as many early models assumed. The upper parts of the scree slopes undergo 'talus' creep which is the aggregate movement caused by rockfall and other impacts, FREEZE–THAW CYCLE or frost heave activity, percolating flows, animals, etc. Loose material on very steep sections may undergo dry avalanching but such failures are usually small. Little if any talus creep seems to occur on coarse basal screes.

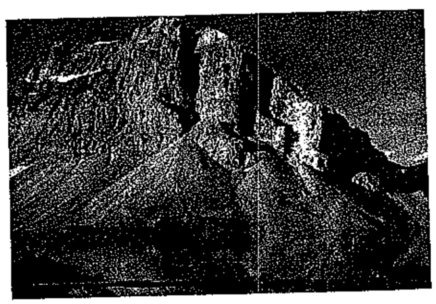


Plate 119 Multiprocess screes, Bow Lake, Alberta, Canada. Complex sheet and scree cones showing the varying influence of alluvial, debris flow and snow avalanche activity on the detailed morphology of these landforms predominantly created by rockfall. The basal area of these cones show evidence of permafrost creep and incipient (or arrested) rock glacier forms

Screes are most frequently studied in periglacial environments. At some of these sites a permanent snow (firn) patch may develop at the base of the slope. Debris landing on this icy surface slides to the base accumulating as a ridge (PROTALUS RAMPART or nivation ridge). In alpine areas, thick talus accumulations may develop PERMAFROST. Subsequent deformation and creep of this rock/ice mixture leads to ROCK GLACIER development.

References

Carson, M.A. (1977) Angles of repose, angles of shearing resistance and angles of talus slopes, *Earth Surface Processes* 2, 363–380.

Church, M., Stock, R.F. and Ryder, J.M. (1979) Contemporary sedimentary environments on Baffin Island, N.W.T., Canada: debris slope accumulations, *Arctic and Alpine Research* 11, 371–402.

Further reading

Luckman, B.H. (1988) Debris accumulation patterns on talus slopes in Surprise Valley, Alberta, Canada, *Géographie physique et Quaternaire* 42, 247–278.

Rapp, A. and Fairbridge, R.W. (1968) Talus fan or cone: scree and cliff debris, in R.W. Fairbridge (ed.) *The Encyclopaedia of Geomorphology*, 1,107–1,109, New York: Reinhold.

SEE ALSO: frost and frost weathering; geomorphological hazard; grèze litée; hillslope, form; hillslope, process; mass movement

BRIAN LUCKMAN

SEA LEVEL

Sea level is the divide between the marine and terrestrial realms; above it, the world is dominated by air, erosion and creatures that contend with gravity; below lies a submerged world dominated by sedimentation and neutrally buoyant animals. Relative to the landmasses, the position of sea level has fluctuated throughout the geologic past, owing to changes in both the quantity of ocean water and the geometry of the ocean basins. Although the total quantity of water in the hydrosphere probably has changed little since Archean times, the fractions held in land reservoirs – glaciers, lakes, ground water and, in particular, continental ice sheets – have fluctuated significantly. For example, if the present Antarctic ice sheet were to melt, sea level would rise by about 55 m; at the height of the last ice age, which is only one of many such events to have punctuated Earth history, sea level was about 120–130 m below its present position.

The proportions of land and sea are basically determined by the fact that continents, being composed of rock lighter than oceanic crust, stand about 4.5 kilometres above the ocean floor (in contrast, hot and oceanless Venus has no features resembling great ocean basins). However, owing largely to slow variations in the rates of seafloor spreading and plate tectonics, the average depth of the ocean basins has varied throughout geological time and shorelines have periodically advanced and retreated across the continental shelves. Figure 142 illustrates examples of observed sea level change on different time scales, from about a year to 10^8 years. On the longer timescale, sea level changed globally with amplitudes up to several hundred metres, largely owing to plate-tectonic changes in ocean basin geometry (Figure 142a). On timescales of tens to hundreds of thousands of years, periodic exchanges of mass between the ice sheets and oceans caused sea-level changes of tens to over a hundred metres in amplitude (Figure 142b).

Global changes of sea level caused by changes of the volumes of seawater or the ocean basins are referred to as *eustatic*. Superimposed on these global signals are more regional and local changes. At decadal, annual and shorter intervals, meteorology – and tide-driven changes become important and vary from place to place (Figure 142c). Over longer times, the relative positions of land and sea are affected by uplift or subsidence of the coastal zone. Observations vary substantially from site to site, even over relatively short distances such as in Scandinavia, where sea level at Ångerman has fallen nearly 200 m in the past 9,000 years while at Andøya the level 9,000 years ago was near its present position. In southern England, levels have risen slowly over the past 7,000 years, but along the Australian margin they have fallen by a few metres during the same interval (Figure 143).

Many factors that contribute to changes in sea level are linked. When ice sheets melt, the resulting sea-level change is spatially variable because the Earth's surface deforms under the changing ice and water load, and because the gravitational potential of the Earth–ocean–ice system also changes. The combined deformation–gravitational effects are referred to as the *glacio-hydro-isostatic* contributions to sea level, and it is they that cause the spatial variability illustrated in Figure 143.

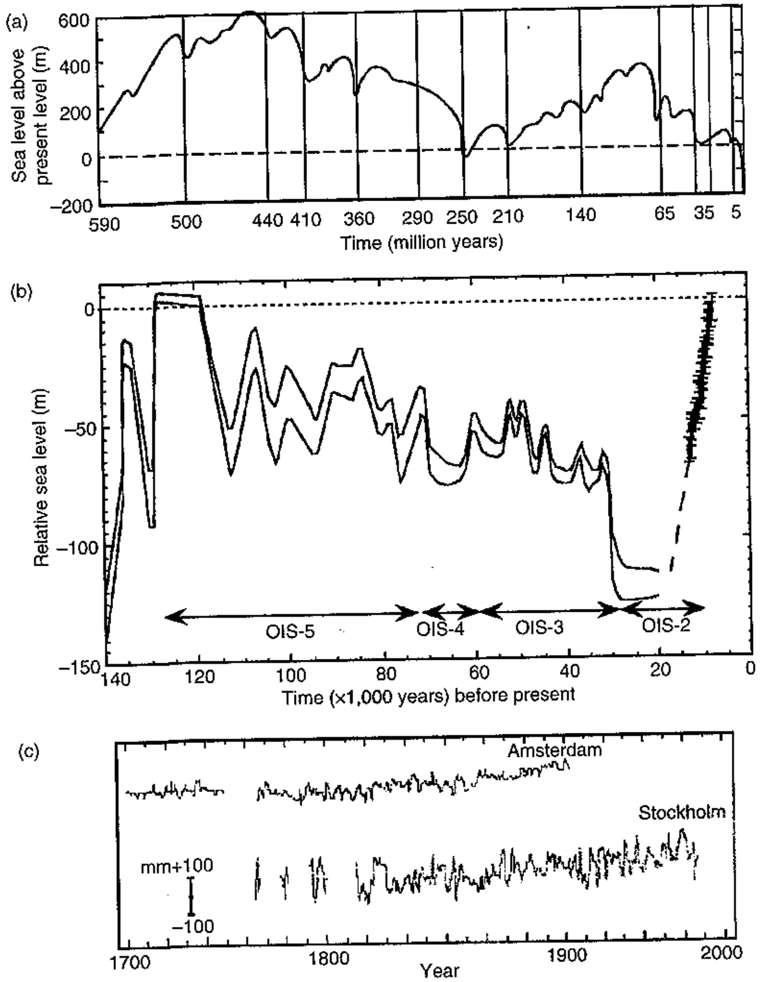


Figure 142 Sea-level variation at three timescales. (a) ~10^8 years, inferred from seismic sequence stratigraphy (Hallam 1992; Haq *et al.* 1988). The higher frequency changes reflect both global and local effects; the large, slow changes reflect continental breakup and changes of ocean ridge systems. (b) ~10^5 years of relative sea level at Huon Peninsula (HP), Papua New Guinea, driven by changes in northern continental ice sheets. Bars show marine oxygen isotope stages (OIS) discussed in text. (c) 10^0–10^2 years, measured by tide gauges from Amsterdam and Stockholm (a secular fall of ~4 mm/year has been removed from the Stockholm record): these changes are primarily of climatic origin and the apparent small rise starting AD ~1880 may reflect global warming (from Lambeck and Chappell 2001)

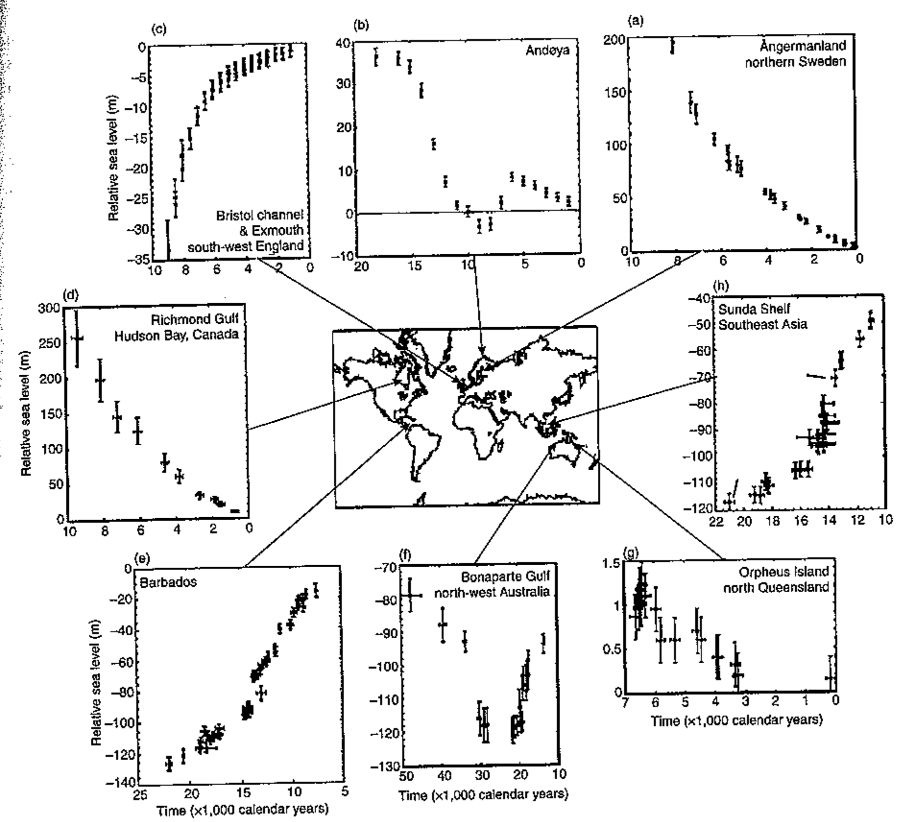


Figure 143 Observed variability of sea-level change in last 20,000 years from tectonically stable areas or sites where the tectonic rate is known and has been removed. (a) Ångerman, Gulf of Bothnia, Sweden. (b) Andøya, Nordland, Norway. (c) South of England. (d) Hudson Bay, Canada. (e) Barbados. (f) Bonaparte Gulf, north-west Australia. (g) Orpheus Island, north Queensland, Australia. (h) Sunda Shelf, Southeast Asia. (Note: scales differ from graph to graph). (Data sources given in Lambeck and Chappell 2001).

Isostatic warping and subsidence also occurs in sedimentary basins, in response to the accumulation over millions of years of sediments many kilometres deep. Furthermore, *tectonic* movements drive mountain uplift, enhancing landscape denudation and leading to rapid accumulation in sedimentary basins, and through coastal uplift or subsidence also contribute to changes of relative sea level.

Knowledge of the complex history of relative changes of sea level has applications in fields as diverse as understanding climate changes, analysing the structure of petroleum-bearing sedimentary basins, and determining deep-earth properties such as the viscosity of the Earth's mantle. Once comprehensive sea-level models are developed, it becomes possible to test hypotheses about the migrations of flora and fauna across shallow seas that are now covered by the ocean. Finally, to understand future sea-level rise under atmospheric greenhouse conditions, the background 'natural' signal must be known. The success of the outcomes of the various sea-level studies depends very much on the ability to separate the different contributions – eustatic, isostatic and tectonic – in the observational record.

Observational evidence

Evidence for historical sea-level changes comes from tidal marks and gauges, usually in ports and harbours, as well as buildings and other structures in littoral towns such as Venice. For the geologic past, the evidence occurs mostly in the form of sediments and biohermal reefs that were formed in coastal and nearshore situations. Upper Quaternary sea-level changes are usually pieced together from raised or submerged shoreline features, including shallow-water coral reefs (Figure 144). Further back in time, relative sea-level changes can be deduced from sedimentary basins, in which eustatic variations are registered in cyclic sediments (cyclothems) (Figure 145), and by alternating transgressive and regressive sediment tracts. Using the methods of *sequence stratigraphy* based on seismic and drillhole or outcrop data, the locus of coastal-zone sediments can be traced throughout a given basin sequence, allowing the course of sea-level changes relative to the basin to be identified (Hallam 1992; Haq *et al.* 1988).

A relative sea-level curve for a given area can be established from age–height data for a series of ancient shorelines, or other *indicator deposits* such as shallow marine sediments for which the depth of deposition relative to sea level is determinable. Terrestrial deposits within a sequence, such as peats and floodplain sediments, may usefully indicate levels not reached by the sea. Various methods are used to establish the ages. For deposits less than ~40,000 years old, *radiocarbon* dating is used widely, although *uranium-series* dating is preferable in the case of coral formations and has a greater time range, extending to ~0.5 Myr. Thermoluminescence (*TL*) and optically stimulated luminescence (*OSL*) methods are increasingly used for dating Upper Quaternary shoreline deposits, and amino acid racemization has proved to be useful despite its low precision. Sea-level indicator deposits also are correlated to marine oxygen isotope records, described later, that have a chronology based on slow variations in the Earth's orbit, which affect the seasonal receipts of solar radiation and acted as an ice-age pacemaker. *Orbital chronology* rests on astronomical observations and has been extrapolated several million years into the past. Finally, the dating of older deposits and sedimentary basin sequences generally rests on *geomagnetic reversal*

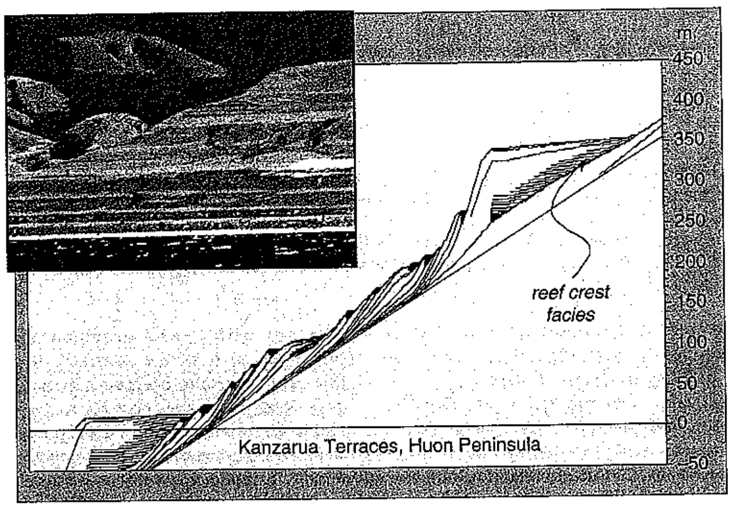


Figure 144 Typical expression of Late Quaternary sea-level changes superimposed on tectonically rising terrain: coral terraces Huon Peninsula, Papua New Guinea. The top terrace at right of the inset photo is 350 m above sea level and formed at the beginning of the Last Interglacial, ~127,000 years ago. Each of the smaller reef structures within the downstepping stratigraphic sequence was formed during a sea-level oscillation within the last glacial cycle. Subtracting the effect of uplift yields the sea-level curve shown at Figure 142b

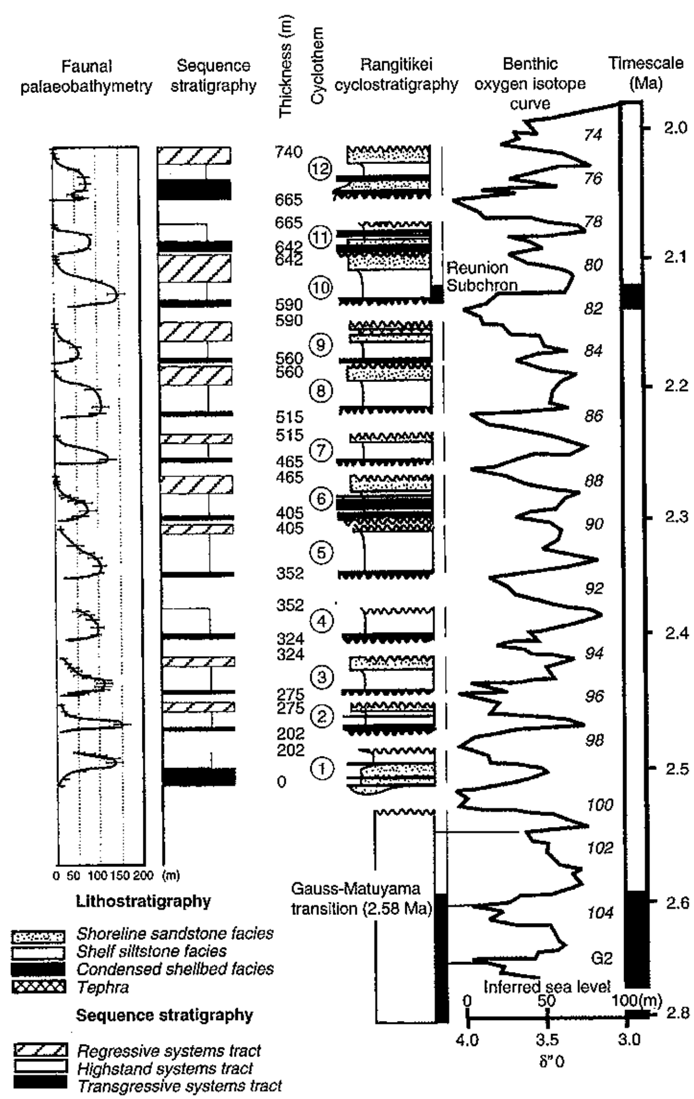


Figure 145 Late Pliocene sea-level changes inferred from shallow marine cyclothems in New Zealand. Centre columns show 12 sedimentary cycles; left column shows cyclic variations of water depth at the site of sedimentation, inferred from fossil marine invertebrates. Correlations to global marine oxygen isotope cycles (Shackleton *et al.* 1995) is shown at right, tied to magnetic reversal chronology (from Pillans *et al.* 1998)

chronology or on microfossil-based correlations with standard stratigraphic sequences tied to this chronology, which in turn is secured by potassium-argon dating methods.

Separation of eustatic, isostatic and tectonic components

RAW SEPARATION OF UNIFORM UPLIFT OR SUBSIDENCE

A first step towards separation of the tectonic, isostatic and eustatic contributions is subtraction of any obvious vertical crustal movements from a relative sea-level curve. More often than not, this is done by assuming that the sea level for some reference deposit in the record is known and also that the rate of uplift (or subsidence) was constant throughout the duration of the record. With these assumptions, the local sea level S represented by a deposit of age t that accumulated at depth d below sea level and now is height H above present sea level, is given by

$$S = H + d - Ut \quad \text{with} \quad U = (S_r - H_r + d_r)/t_r \qquad (1)$$

where H_r, d_r and t_r are height, depositional depth and age of the reference deposit, and S_r is the sea level at its time of formation. For Upper Quaternary studies, the local height of the Last Interglacial shoreline is widely used to determine uplift rate U, as evidence reviewed below suggests that sea level then was little different from that of today. However, for every study, each variable in (1) has an error term arising from dating errors and uncertainties in field relationships. Moreover, in assuming a constant rate of uplift, this approach neglects reversing vertical movements that arise from the global glacio-hydro-isostatic response to advancing and retreating icesheets. Figure 142b shows the 'uplift-free' sea-level curve derived by this method from the coral terraces illustrated at Figure 144.

Similar principles are used to derive sea-level changes from sedimentary basin sequences, although here the vertical movement is downwards. In terms of sedimentary facies, individual cyclothems often are very similar from bottom to top of a thick cyclothemic sequence (Figure 145), implying fairly uniform subsidence of the basin.

GLACIO-HYDRO-ISOSTASY

When ice sheets melt, to a first approximation the sea level rises by an amount $\Delta\zeta_e(t)$ related to the land-based ice volume V_i (using the notation of K. Lambeck: see Lambeck and Chappell 2001),

$$\Delta\zeta_e(t) = -(\rho_i/\rho_o)\int(1/A_o(t)\ dV_i/dt)\,dt \qquad (2)$$

where $A_o(t)$ is the ocean surface area (which changes as sea level rises or falls) and ρ_i, ρ_o are the average densities of ice and ocean water, respectively. $\Delta\zeta_e(t)$ is the ice-volume equivalent sea-level change (or *equivalent sea-level change*), which equals eustatic change if no other factors contribute to changes in ocean volume. The relative sea-level change $\Delta\zeta_{rsl}(\varphi, t)$ at position and time t, ignoring tectonic displacements, is

$$\Delta\zeta_{rsl}(\varphi, t) = \Delta\zeta_e(t) + \Delta\zeta_i(\varphi, t) + \Delta\zeta_w(\varphi, t) \quad (3)$$

where $\Delta\zeta_i$ and $\Delta\zeta_w$ are the glacio- and hydro-isostatic contributions. Both are functions of position and time. The water depth or terrain height, expressed relative to coeval sea level, is

$$h(\varphi, t) = h(\varphi, t_0) - \Delta\zeta_{rsl}(\varphi, t) \qquad (4)$$

where $h(t_0)$ is the present-day (t_0) bathymetry or topography at φ. Both isostatic terms in (3) are functions of Earth rheology as well as of fluctuations in the ice sheets over time.

In formerly glaciated areas, the glacio-isostatic term $\Delta\zeta_i(\varphi, t)$; dominates during and after deglaciation, and leads to uplift at a rate that can exceed the global eustatic rise, so that sea level locally falls (the Ångerman result: Figure 143). Rebound is smaller near the ice margin and although it may dominate initially, the global sea-level rise becomes important later. When all melting has ceased, the residual rebound leads to falling local sea level (the Andøya result: Figure 143). During ice sheet growth, mantle material beneath the loaded area is displaced outward and broad bulges develop around the perimeter, which subside when the ice melts, leading to slowly rising local sea level after melting has ceased. Much further from the ice, the water load becomes the dominant cause of planetary deformation (the hydro-isostatic contribution $\Delta\zeta_w(\varphi, t)$), producing subsidence of the seafloor and adjacent margins. The effect is most pronounced at continental margins far from the ice sheets, such as the Australian coast (Figure 143), and once melting has ceased, sea levels continue to fall at a slow but perceptible rate. The amplitude of this postglacial 'highstand' effect can vary by several metres from site to site.

Isostatic corrections and a global eustatic curve can be derived from local sea-level curves, using a rheologically appropriate Earth model to predict surface deformation in response to changing ice and water loads, with the ocean surface remaining a gravitational equipotential surface at all times. The sea-level signal at sites far from the former ice margins approximates the equivalent sea-level function to about 10–15 per cent and the isostatic contribution is mainly from water loading, which is insensitive to the details of the ice sheets, provided that the total ice volumes are correct to within about 10 to 20 per cent. Hence, through an iterative procedure, it becomes possible to estimate changes in ice volumes V_i from observed sea-level changes $\Delta\zeta_{rsl\ (obs)}$, using (2), and (5), below:

$$\Delta\zeta_{rsl} = \Delta\zeta_{rsl\ (obs)} - (\Delta\zeta_i + \Delta\zeta_w) \qquad (5)$$

The use of local sea-level curves from widely separated places allows Earth rheology parameters and models of ice distribution to be evaluated. Recent models include deformation of the basins over time, movement of grounded ice across the shelves, modification of sea level by the time-dependent gravitational attraction between the solid Earth, ocean, and ice, and the effect of glacially induced changes in Earth rotation on sea level.

Sea level through the last glacial cycle

Sea-level data for the last glacial cycle are more plentiful than for earlier periods and, at Huon Peninsula, Papua New Guinea, provide a near-complete relative sea-level curve (Figure 142b), which has been used for reconstructing ice-equivalent sea level for the past 140,000 years (Lambeck and Chappell 2001). Results indicate that ice melting has varied since the Last Glacial Maximum (LGM), with two periods of rapid sea-level rise from ~16,000 to 12,500 and from 11,500 to 8,000 years ago, separated by the Younger Dryas (YD) cold episode when sea level seems to have risen less rapidly. By 7,000 years ago, the northern ice sheets except for Greenland had gone and ocean volume approached its present level, but Antarctic ice melting may have since contributed a few metres of equivalent sea level.

The Last Interglacial, when sea level was similar to the present, ended about 118,000 years ago with rapidly falling sea level, associated with the growth of northern continental ice. Cyclic sea-level changes from 118,000 to ~60,000 years reflect the effect of the 20,000-year orbital precession cycle on the ice sheets, although other fluctuations also appear. These occur repeatedly about every 6,000 years from ~60,000 to 30,000 years, with amplitudes of 10–15 m, and each rise apparently coincides with a major episode of ice-rafted sediment deposition recorded in the North Atlantic, suggesting that the rise was caused by a large, rapid discharge of continental ice.

Oxygen isotopes and long sea-level records

Long, continuous records of oxygen isotopes in calcareous foraminifera preserved in deep-sea sediments have become standard records of Quaternary sea-level and temperature changes. The oxygen isotope ratio in foraminifera, conventionally expressed as $\delta^{18}O$ (the ‰ difference of $^{18}O/^{16}O$ in a sample from an international standard), depends on the temperature and isotope ratio of the seawater in which they lived. Furthermore, the seawater isotopic ratio is related to the size of polar ice sheets, because ^{18}O is preferentially removed from atmospheric water vapour as it makes its way poleward, causing the icecaps to be depleted in ^{18}O relative to seawater by ~25–55 ‰, varying with atmospheric temperature. Thus, foraminiferal isotopes $\Delta\delta^{18}O_f$ respond to ice volume changes ΔV_i according to

$$\Delta\delta^{18}O_f \approx \delta^{18}O_i\ \Delta V_i/V_w + C_T\,\Delta T + \varepsilon \qquad (6)$$

where ΔV_i is relative to present day V_i, V_w is the present volume of seawater (mean $\delta^{18}O$ of modern seawater $= 0$ ‰$_{SMOW}$), C_T is the coefficient of temperature-dependent isotope fractionation for calcite (-0.23 ‰ $°C^{-1}$), ΔT is temperature change, and ε represents any local change of seawater $\delta^{18}O$ not related to ΔV_i. (Equation 6 is approximate because the mean isotopic composition of the ice sheets $\delta^{18}O_i$ is assumed constant, but the uncertainty here probably is smaller than the effects of ice volume and temperature.) Finally, the first term on the RHS of (6) can be expressed in terms of equivalent sea level $\Delta\zeta_e$: comparison of isotopes and sea levels for the last glacial cycle indicates that $\Delta\delta^{18}O_w/\Delta\zeta_e \sim -0.009$ ‰ m^{-1}.

The numerical value of $\delta^{18}O_f$ becomes increasingly positive under a fall of both sea level and temperature, which typically happens under a major ice-sheet advance. Hence, marine oxygen isotopes provide composite records of sea level and temperature: for example, Figure 145 illustrates typically close correspondence between sea level and isotopic cycles. In a longer time frame, Tertiary isotope records reveal both progressive cooling and the onset of ice-driven sea-level cycles

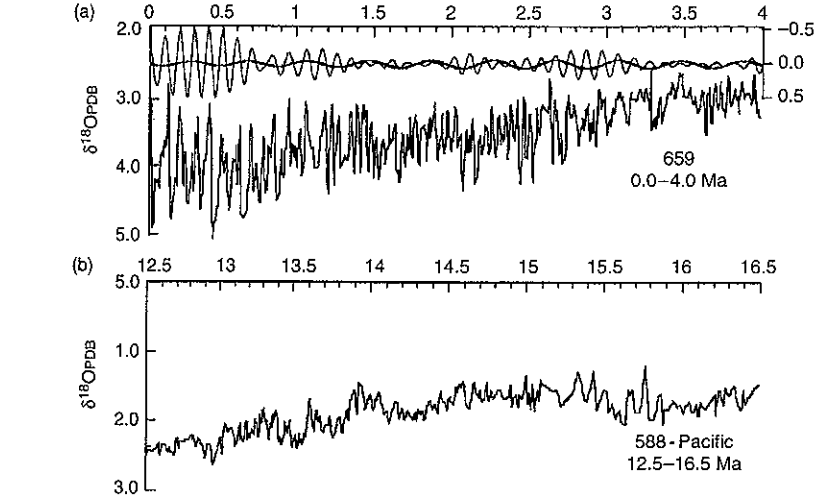


Figure 146 Contrasting records of marine oxygen isotopes from 0–4 Myr (Plio-Pleistocene) and 12.5–16.5 Myr (Miocene). The increase of amplitude in $\delta^{18}O$ cycles around 2.6 Myr represents the onset of large fluctuations of ice sheets and sea level, which become even more pronounced around 1 Myr, when they adopt a characteristic period of ~100,000 years. The upper curve in (a) shows the changing amplitude of the 100,000 year cycle (after Zachos *et al.* 2001)

in the Pliocene (Figure 146). As the history of ocean temperature becomes increasingly well determined, through trace-element analysis of microfossils and other techniques, the sea-level contribution in such records is now being isolated (see Zachos *et al.* 2001).

Shoreline reconstructions

Once a global eustatic curve $\zeta_e(t)$ is established, the course of shoreline changes through time can be predicted. Provided that the present-day shallow water bathymetry is known with high resolution, water depths for any region at any time within the range of the eustatic curve follow from (4), and the palaeo-shorelines at time t correspond to the contours $h(\varphi, t) = 0$. Thus, it becomes possible to examine the migrations of shorelines through time for intervals for which sufficient observational data exist to constrain the isostatic variables. Predictions of shorelines since the time of the LGM have been published for both global and regional reconstructions, which can provide useful insights into the interpretation of prehistoric sites. As Lambeck (1996) has shown, for example, the interpretation of post-Palaeolithic archaeology of the Aegean is intimately linked to reconstructions of shoreline changes, which were a powerful factor in trade and changing human activities in the region.

Geomorphic consequences of sea-level changes

Rising sea level at the end of the Pleistocene Period led to widespread geomorphologic changes in coastal regions. Under the influence of ice-age low sea levels, today's coastal valleys tended to become incised well below present sea level, only to turn into traps for floodplain aggradation when sea level rose, as ice sheets retreated. The alternation of incision and aggradation doubtless was repeated in each glacial cycle, throughout the Quaternary, leading to development of broad floodplains on thick sediment valley-fills, which in tectonically stable regions often extend hundreds of kilometres inland. Near-coastal limestone regions under the influence of low sea levels developed vadose karst systems below present sea level, including stream passages and speleothem formations. At coastlines, a range of landforms were the result of the postglacial sea-level rise, from drowned valleys through sand-barrier basins to estuarine and deltaic plains, the particular form depending on sediment supply, tidal range and wave energy. In tropical seas, coral reefs re-established after being reduced by subaerial erosion during glacial-age lowstands.

In tectonically stable terrain, the geomorphic expression of any given Quaternary sea-level cycle tends to overprint the record of earlier cycles. In contrast, tectonic uplift results in flights of coastal terraces built at times of high relative sea level (e.g. Figure 144) that often pass inland into flights of river terraces. However, even though a terraced river valley may grade to present sea-level, not all its terraces necessarily relate to sea level highstands. Where continental shelves are broad, a river carrying sufficient fluvial sediment, when sea level was low and the coast far seawards of its present position, may have aggraded to a level higher than the recent floodplain, which thus becomes inset below a Pleistocene lowstand terrace. Given the tendency for all such features to become degraded and overprinted, the geomorphic legacy of past sea levels – while very sharp for recent events – becomes increasingly blurred with time. In contrast, the sedimentary record remains relatively intact.

References

Hallam, A.J. (1992) *Phanerozoic Sea-level Changes*, New York: Columbia University Press.
Haq, B., Hardenbol, J., Vail, P.R., Hardenbol, J. and Baum, G.R. (1988) Sea-level change: an integrated approach, *Society of Economic Paleontologist and Mineralogists Special Publication* 42, 71–108.
Lambeck, K. (1996) Sea-level change and shore-line evolution in Aegean Greece since Upper Palaeolithic time, *Antiquity* 70, 588–611.
Lambeck, K. and Chappell, J. (2001) Sea level change through the last glacial cycle, *Science* 292, 679–686.
Pillans, B., Chappell, J. and Naish, T. (1998) A review of the Milankovitch beat: template for Plio-Pleistocene sea-level changes and sequence stratigraphy, *Sedimentary Geology* 122, 5–21.
Shackleton, N.J., An, Z., Dodonov, A.E., Gavin, J., Kukla, G.J., Ranov, V.A. and Zhou, L.P. (1995) Accumulation rates of loess in Tadjikistan and China: relationship with global ice volume cycles, *Quaternary Proceedings* 4, 1–6.
Zachos, J., Pagani, M., Sloan, L., Thomas, E. and Billups, K. (2001) Trends, rhythms, and aberrations in Global Climate 65 Ma to Present, *Science* 292, 686–693.

JOHN CHAPPELL

SEAFLOOR SPREADING

Seafloor spreading is the process by which new oceanic crust is generated at mid-ocean ridges, the long linear belts of elevated seafloor that lie at the centres of most ocean basins (see SUBMARINE LANDSLIDE GEOMORPHOLOGY). The term was coined by Dietz (1961) for the idea (jointly proposed with Hess 1962) that new crust is formed by magmatic intrusion along the crests of mid-ocean ridges, and that conveyor belts of crust thus formed move symmetrically away from the ridges, driven by convection currents in the underlying mantle. Their suggestion provided the first viable mechanism by which continental drift might take place. Acceptance of the concept came only after the demonstration by Vine and Matthews (1963) that linear magnetic anomalies, which are symmetrical on either side of a mid-ocean ridge, record reversals of the Earth's magnetic field as newly formed crust cools and moves laterally away from the axis. The seafloor spreading hypothesis was the direct catalyst for the development of the concept of PLATE TECTONICS, the grand unifying theory of Earth sciences. This posits that the Earth is capped by a number of rigid plates, containing both continental and oceanic crust, that deform only at plate boundaries.

The large-scale motion of the Earth's plates is now believed to be driven primarily by the pull of dense slabs of oceanic lithosphere as they are subducted into the mantle at convergent margins (see WILSON CYCLE). Seafloor spreading at divergent oceanic plate boundaries may therefore be regarded as an essentially passive process: plates are pulled rather than pushed apart. Separation of the oceanic plates induces upwelling of the convecting, plastically deforming mantle asthenosphere. As it rises up and decompresses the mantle peridotite starts to melt. This generates a basaltic liquid which separates from its host and rises upward to feed a magma chamber beneath the ridge axis, thence solidifying to form ocean crust.

The rate at which seafloor spreading occurs varies from place to place along the 60,000 km long global mid-ocean ridge system, from less than 1 cm yr^{-1} at the Gakkel Ridge (in the Arctic Ocean) and parts of the Southwest Indian Ridge, to 16 cm yr^{-1} at the East Pacific Rise (west of Peru). The oldest surviving ocean floor lies in the north-west Pacific Ocean and is of Jurassic age (approximately 180 Ma).

Seismic refraction experiments indicate that the ocean crust is generally 6–7 km thick and has a layered internal structure. A seismic discontinuity (the Mohorovicic discontinuity or 'Moho') separates it from rocks with typical mantle velocities below. The seismic layering has been correlated directly with the lithological layering observed in ophiolites, which are regarded as on-land fragments of ocean crust and shallow mantle. This has led to the conventional view of ocean crustal structure as a simple, uniform 'layer-cake' internal structure composed (from top to bottom) of: basaltic lavas, usually with lobate, pillowed morphology indicative of submarine eruption; a parallel 'sheeted' swarm of dolerite dykes that fed the lava flows; and coarse-grained gabbros and ultramafic (olivine-rich) plutonic rocks that crystallized slowly in the magma chamber.

In early models for the generation of ocean crust these magma reservoirs were envisaged as huge bodies the height of the entire lower crust and tens of kilometres wide beneath spreading axes. However, seismic reflection experiments at the fast-spreading East Pacific Rise in the late 1980s showed that no such magma body exists: purely molten material is instead restricted to a thin lens (probably of the order of 100 m thick) in the middle crust, overlying a much larger region in the lower crust that appears to be a partially molten mush composed of crystals with small amounts of interstitial melt. At slower spreading rifts such as the Mid-Atlantic Ridge even this thin magma lens is normally absent, implying that the lower crust there probably freezes solid in between melt delivery events from the mantle, and that magma supply in these environments may be relatively reduced (e.g. Sinton and Detrick 1992).

Ridges may be offset by several hundred kilometres by transform faults, across which lateral motion occurs as seafloor spreads in opposite directions on either side. Smaller scale offsets or discontinuities of the spreading axis between transform faults define the boundaries between individual spreading segments, which may be regarded as elongate volcanoes more or less aligned along the ridge crest (Macdonald *et al.* 1988).

The morphology of the region around the ridge crest is very much dependent upon spreading rate. Whereas fast-spreading ridges are characterized by relatively smooth seafloor with an elevated ridge crest, slower spreading ridges (less than ~5 cm yr^{-1}) are instead marked by a very rough seafloor and an axial valley that may be more than 2 km deeper than the surrounding walls. Mantle peridotite, now altered by the action of seawater to serpentinite, is commonly recovered in these axial valleys. The rough seafloor results from extensional faulting, which helps to accommodate separation of the plates when magma supply to the ridge is low. Some of the fault planes have very low dip angles and can accommodate tens of kilometres (in excess of a million years' worth) of displacement on a single structure. These structures, termed 'detachment faults', provide a mechanism by which mantle and lower crustal rocks may be exhumed onto the seafloor. Mantle rocks may also be exposed at some slow-spreading ridges because magma supply to the ridge axis was so low that a continuous layer-cake magmatic crustal layer was never produced in the first place (Cannat 1993). In places, therefore, the seismically defined crustal layer may be composed partly or completely of serpentinite, which has a velocity much lower than fresh peridotite but similar to that of basalt or gabbro. Ocean crustal structure, particularly at slow-spreading ridges, is now understood to be far more heterogeneous than originally suggested on the basis of the seismic refraction experiments. Fast-spread ocean crust may have a more regular 'layer-cake' architecture.

Total crustal production by seafloor spreading at mid-ocean ridges is estimated at 18 $km^3 yr^{-1}$, generating a thermal flux equivalent to ~50 megawatts per kilometre of spreading ridge worldwide. This heat is extracted from the newly formed crust primarily by hydrothermal convection: seawater descends through cracks in the crust, heats up and is eventually vented back into the water column as 'black smoker' fluid. This fluid is hot (up to 400 °C) and rich in metals that have been stripped from the basaltic crust. Volumes are such that the equivalent of the entire world ocean is believed to circulate through the crust every seven million years or so. Hydrothermal circulation therefore plays an essential role in regulating the composition of seawater on geological timescales.

Fluid circulation and cooling persists away from the ridge crests. It causes the lithosphere – the mechanically rigid plate – to increase progressively in thickness as it moves away from the ridge. As the uppermost mantle cools below ~1,000 °C it ceases to be able to flow in the convecting asthenosphere and moves with the overlying crust, in effect being attached or welded to the base of the ocean crust. Mature ocean lithosphere is up to 100 km thick, all but the uppermost few kilometres being rigid mantle.

On a broad scale the ridge crest and surrounding seafloor is elevated relative to the abyssal plains because of the lower density of hot, partially molten asthenospheric mantle flowing upward beneath the ridge crest. Moving the lithosphere away from the region of upwelling and thickening it causes the seafloor to subside (see ISOSTASY). The rate of subsidence is proportional to the square root of the age of the crust and is independent of spreading rate, explaining why the width of the elevated region around the East Pacific Rise is far broader than that around the Mid-Atlantic Ridge.

References

Cannat, M. (1993) Emplacement of mantle rocks in the seafloor at mid-oceanic ridges, *Journal of Geophysical Research* 98, 4,163–4,172.

Dietz, R.S. (1961) Continent and ocean basin evolution by the spreading of the sea floor, *Nature* 190, 854–857.

Hess, H.H. (1962) History of the ocean basins, in A.E.J. Engel, H.L. James and B.F. Leonard (eds) *Petrologic Studies: A Volume to Honor A.F. Buddington*, Denver: Geological Society of America.

Macdonald, K.C., Fox, P.J., Perram, L.J., Eisen, M.F., Haymon, R.M., Miller, S.P., *et al.* (1988) A new view of the mid-ocean ridge from the behaviour of ridge-axis discontinuities, *Nature* 335, 217–225.

Sinton, J.M. and Detrick, R.S. (1992) Mid-ocean ridge magma chambers, *Journal of Geophysical Research* 97, 197–216.

Vine, F.J. and Matthews, D.H. (1963) Magnetic anomalies over ocean ridges, *Nature* 99, 947–949.

Further reading

Kearey, P. and Vine, F.J. (1996) *Global Tectonics*, 2nd Edition, Oxford: Blackwell.

Nicolas, A. (1995) *The Mid-Oceanic Ridges: Mountains Below Sea Level*, Berlin: Springer-Verlag.

Open University Course Team (1998) *The Ocean Basins: Their Structure and Evolution*, 2nd Edition, Oxford: Butterworth-Heinemann.

Oreskes, N. and Le Grand, H. (eds) (2001) *Plate Tectonics: An Insider's History of the Modern Theory of the Earth*, Boulder, CO: Westview Press.

CHRIS MACLEOD

SEDIMENT BUDGET

A sediment budget is a quantitative accounting of the rates of production, transport and discharge of detritus in a geomorphic system such as a DRAINAGE BASIN, coastal BEACH, offshore zone, hillslope, river channel, GLACIER, or any landscape unit around which boundaries can be drawn. Sediment budgets apply the principle of conservation of mass to geomorphic systems, an approach that became popular in the 1970s. Sediment budgets provide a tool for research geomorphologists to judge the relative importance of sediment sources, storage sites and transfer processes, including how they change over time. Sediment budgets are also useful tools for resource management when it becomes important to distinguish human impacts on geomorphic systems from those that would have occurred without human interference (Reid and Dunne 1996). Studies have been reported that use sediment budgets to document the effect of agriculture, forestry, road construction, urbanization, DAMs, wildfires and mining. Sediment budgets have been used to construct a more complete picture of the distribution of sediment sources within a drainage basin (e.g. Marston and Dolan 1999) and the heavy metals that can be associated with the sediment (e.g. Marcus *et al.* 1993).

Four basic steps must be followed to construct a sediment budget (Lehre 1982): (1) define the boundaries of the geomorphic system; (2) identify the processes and sites of EROSION, transport and storage (deposition) in the geomorphic system, including the linkages between them; (3) quantify the contribution of each over space and time; and (4) set up an accounting sheet that balances the sediment production, sediment yield (see SEDIMENT LOAD AND YIELD), and storage. The first step depends on the ability to recognize the boundaries of the geomorphic system to be studied. Most sediment budgets have been prepared for drainage basins and sandy beaches, both of which have readily recognized boundaries. However, sediment budgets have also been attempted for KARST and glacier systems, where subsurface passages transport significant amounts of sediment that is difficult to trace and measure. Sediment budgets generated by wind in the form of sandstorms and DUST STORMs have been rare, although this component has been used to explain the discrepancies between inputs, change of storage and outputs – a dangerous practice unless all transfer and storage components have been measured with great accuracy (Hill *et al.* 1998). The second step requires training, experience and expertise with the full range of field, lab and office techniques in geomorphology (e.g. aerial photography and historical maps) to discern evidence of the range of possible processes and storage sites. A flow chart may help to visualize the processes, storage sites, outputs of sediment and linkages between them (Figure 147). The third step requires measurement of these processes using an appropriate choice of techniques. Direct measurement of processes is beset by problems, although controlled field experiments can help. In some cases, ^{137}Cs measurements have been used in combination with more traditional techniques to compile tracer-based sediment budgets (e.g. Walling *et al.* 2002). Geomorphologists might also use predictive equations, examine soil profiles, undertake photogrammetric measurements from aerial photographs, or conduct lab measurements of soil and sediment characteristics. Changes in sediment storage can be derived from geophysical surveys, morphometric modelling or field surveys using anchor chains.

One of the classic sediment budget studies in understanding the importance of sediment storage was undertaken by Trimble (1983). He compiled a sediment budget for the 360 km² Coon Creek watershed in Wisconsin for two periods: 1853–1938 when soil erosion was severe because of poor land management, and 1938–1975 following the implementation of SOIL CONSERVATION practices. Sediment yield at the mouth of the WATERSHED was essentially identical for the two periods, from which one might conclude that soil conservation practices failed to have the desired effect. However, when one examines the sediment budget data, it becomes apparent that upland erosion had been reduced by 26 per cent. Sediment yield figures remained high even after soil conservation measures had been implemented because sediment that had been stored in tributary valleys and the upper main valley during the early period had been mobilized during the latter period. This removal from storage offset the reduction in hillslope erosion, with the net effect that overall sediment yields remained essentially unchanged. One of the main lessons of erosion studies has been learned from Trimble's study: sediment yield values alone are not a good indicator of spatial and temporal changes in erosion within a watershed.

Reneau and Dietrich (1991) used a sediment budget approach in the southern Oregon Coast Range to demonstrate that a steady state condition exists between sediment production from hillslopes and sediment yield from the watershed. The authors suggest that erosion rates were spatially uniform and perhaps even diagnostic of a landscape in geomorphic equilibrium. In earlier work elsewhere in the Oregon Coast Range, Dietrich and Dunne (1978) found that the rate that bedrock in the Rock Creek watershed was converted to soil, with the associated increase in volume by a factor of 1.17, was equivalent to the rate of DENUDATION calculated for the watershed over the same time interval. Thus, they conclude that reduction in watershed relief by denudation was balanced by the increase in volume of hillside materials; relief is not changing over time, without accounting for tectonics.

Graf (1994) compiled a fluvial sediment budget for the northern Rio Grande drainage system in New Mexico for the express purpose of tracking deposits likely to contain plutonium. Early development of nuclear weapons at Los Alamos National Laboratory had led to discharge of highly concentrated plutonium wastes to arroyos that had carried the plutonium adsorbed onto sediment into one of the major waterways of western North America. Graf coupled extensive and detailed field mapping with simulation modelling of plutonium movement to demonstrate the importance of large FLOODS in delivering pulses of plutonium downriver. Fortunately, the concentrations of plutonium in stored sediments are generally low, but do vary considerably. With large floods, concentrations will decline over time to approximate levels from atmospheric fallout; without large floods, the portion of plutonium that is associated with bedload will persist. Graf was also able to demonstrate the significance of plutonium storage behind Cochiti Reservoir along the Rio Grande, a site that is now closed for fishing and swimming.

Madej (1987) showed how the residence time of stored sediment could be calculated as part of fluvial sediment budget studies if one knows the sediment storage, expressed as cubic metres per metre length of valley sediment, and the sediment transport rate, expressed in units of cubic metres per year. When the storage is divided by the transport rate, the residence time is derived, in units of years per cubic metre. This strategy was used by Madej to calculate the residence time of sediment in various reaches of Redwood Creek in north-west California. Comparing the residence time to the rate of sediment delivery allows one to determine whether the system is at/near geomorphic equilibrium or likely to experience progressive change over time. If a reach receives

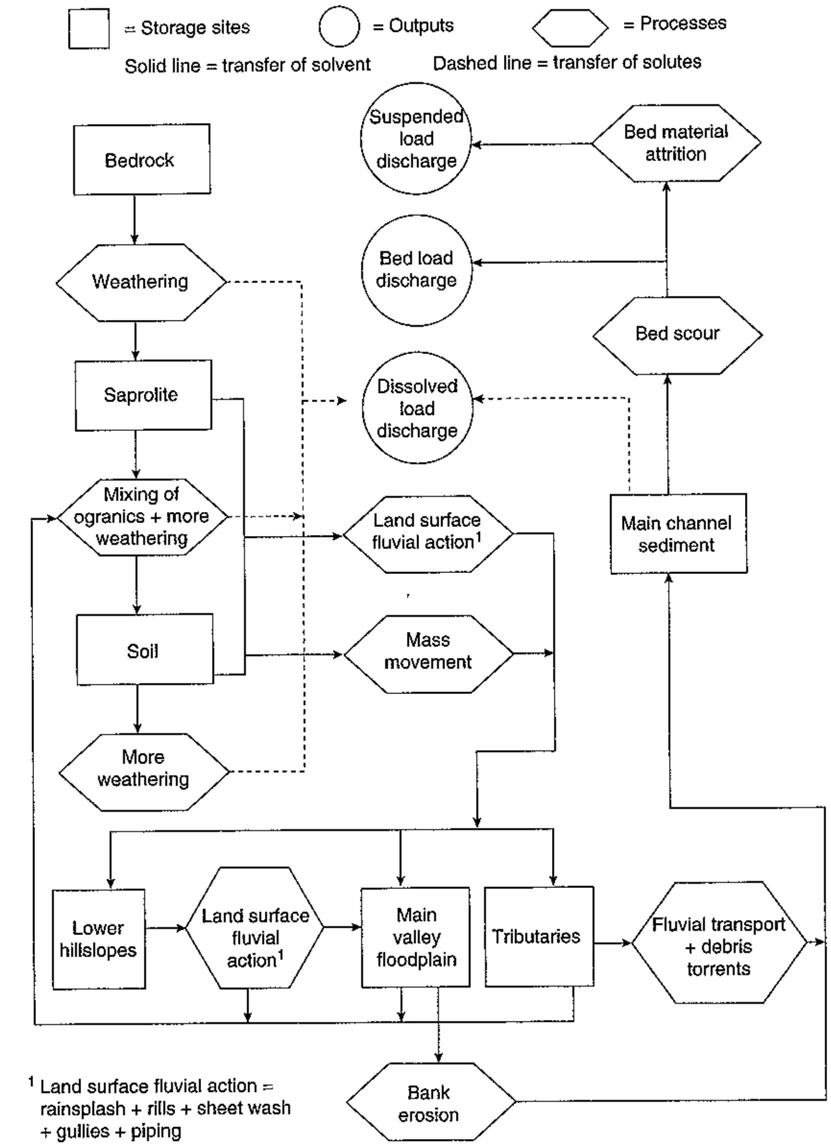


Figure 147 Flow chart for a drainage basin sediment budget (after Dietrich and Dunne 1978)

a volume of sediment from storms with a recurrence interval of fifty years and that sediment is likely to reside in the stream for a hundred years, the stream will experience progressive AGGRADATION. The recovery of Redwood Creek from catastrophic floods and associated MASS MOVEMENT was found to vary depending on the residence time, which in turn varies with landscape position in the valley floor. Indeed, sediment budgets have been used to compare the importance of frequent, low magnitude events with infrequent, catastrophic events in the transport of sediment (e.g. Springer *et al.* 2001).

Sediment budgets have proved especially useful in examining changes in beaches over time. Consider a sediment budget for a sandy beach between two rocky headlands. Sediment sources could include eroding cliffs (see CLIFF, COASTAL), onshore transport, marine erosion of beach material, supply from dunes (see DUNE, COASTAL), subsurface erosion, fluvial input, BEACH NOURISHMENT and LONGSHORE (LITTORAL) DRIFT. Sediment storage could occur on the beach and adjacent inland dunes, as well as in offshore banks, SPITS and bars (see BAR, COASTAL). Sand could be lost through offshore transport, longshore drift, aeolian erosion (see AEOLIAN PROCESSES), and dredging. The effects of various shoreline protection measures can be effectively measured with this approach (Cooper *et al.* 2001). One noteworthy study demonstrated that the loss of sand along beaches of southern California could be attributed to the construction of sediment detention basins in the mountains surrounding the Los Angeles Basin. When it was discovered that valuable recreational beaches were being deprived of river-delivered sand and diminishing in size, artificial nourishment was required at great expense (Cooke 1984).

Sediment budgets are utilized over a wide range of spatial scales, from small plot studies (Duijsings 1987) to continental-scale assessments. New techniques are being developed to construct more accurate sediment budgets. The sediment budget approach is being applied to an ever-expanding variety of geomorphic environments and for discerning the effects of human activities in these settings. Because sediment budgets deal with the sources, storage, throughflow and outputs of sediment in a geomorphic system, they have been characterized as a fundamental method in understanding cascading process systems.

References

Cooke, R.U. (1984) *Geomorphological Hazards in Los Angeles*, London: Allen and Unwin.

Cooper, N.J., Hooke, J.M. and Bray, M.J. (2001) Predicting coastal evolution using a sediment budget approach: a case study from southern England, *Ocean and Coastal Management* 44, 711–728.

Dietrich, W.E. and Dunne, T. (1978) Sediment budget for a small catchment in mountainous terrain, *Zeitschrift für Geomorphologie, Supplementband* 29, 191–206.

Duijsings, J.J.H.M. (1987) *Streambank Contribution to the Sediment Budget of a Forest Stream*, Publication No. 40, Amsterdam: Fysisch-Geografisch en Bodemkundig Laboratorium van de Universiteit van Amsterdam.

Graf, W.L. (1994) *Plutonium and the Rio Grande: Environmental Change and Contamination in the Nuclear Age*, New York: Oxford University Press.

Hill, B.R., Decarlo, E.H., Fuller, C.C. and Wong, M.F. (1998) Using sediment fingerprints to assess sediment-budget errors, North Halawa Valley, Oahu, Hawaii, *Earth Surface Processes and Landforms* 23, 493–508.

Lehre, A.K. (1982) Sediment budget of a small coast range drainage basin in north-central California, in F.J. Swanson, R.J. Janda, T. Dunne and D.N. Swanson (eds) *Sediment Budgets and Routing in Forested Drainage Basins*, General Technical Report PNW-141, 67–77, Portland, OR: USDA Forest Service.

Madej, M.A. (1987) Residence times of channel-stored sediment in Redwood Creek, northwestern California, in R.L. Beschta, T. Blinn, G.E. Grant, G.G. Ice and F.J. Swanson (eds) *Erosion and Sedimentation in the Pacific Rim*, Publication No. 165, Wallingford: International Association of Hydrological Sciences.

Marcus, W.A., Neilsen, C.C. and Cornwell, J. (1993) Sediment budget analysis of heavy metal inputs to a Chesapeake Bay estuary, *Environmental Geology and Water Sciences* 22, 1–9.

Marston, R.A. and Dolan, L.S. (1999) Effectiveness of sediment control structures relative to spatial patterns of upland soil loss in an arid watershed, Wyoming, *Geomorphology* 31, 313–323.

Reid, L.M. and Dunne, T. (1996) *Rapid Evaluation of Sediment Budgets*, Reiskirchen: Catena Verlag.

Reneau, S.L. and Dietrich, W.E. (1991) Erosion rates in the southern Oregon Coast Range: evidence for an equilibrium between hillslope erosion and sediment yield, *Earth Surface Processes and Landforms* 16, 307–322.

Springer, G.S., Dowdy, H.S. and Eaton, L.S. (2001) Sediment budgets for two mountainous basins affected by a catastrophic storm: Blue Ridge Mountains, Virginia, *Geomorphology* 37, 135–148.

Trimble, S.W. (1983) A sediment budget for Coon Creek basin in the Driftless Area, Wisconsin, *American Journal of Science* 283, 454–474.

Walling, D.E., Russell, M.A., Hodgkinson, R.A. and Zhang, Y. (2002) Establishing sediment budgets for two small lowland agricultural catchments in the UK, *Catena* 47, 323–353.

RICHARD A. MARSTON AND MARCUS E. PEARSON

SEDIMENT CELL

A *sediment cell* is a section of the coastal zone where the sediment inputs, throughput and outputs may be considered part of a closed system. Given the vast range of coastal environments and the nature and sources of sediments, a wide range of sediment cells exist, both in form, and temporal and spatial scale. A more general definition is, therefore, a coastal environment where the input, accumulation and output of sediments are part of an interrelated flow of sediments, some of which may be derived from, and/or exported to, adjacent sediment cells.

Komar (1998) discusses sediment cells using the concept of coastal SEDIMENT BUDGETs, based on the example of the Southern California littoral cells. Each cell has a sediment source (river, cliff erosion, etc.), longshore transport, and finally sinks which include submarine canyons, loss to dunes and transport to downdrift cells. Each cell therefore contains a cycle of sediment input, transport and sedimentation, the latter either as accumulation within the cell, or loss onshore, offshore or longshore. He presents a more generalized budget of littoral sediments which includes sediment credit, debit and a balance, the latter resulting in positive beach deposition or negative beach erosion.

Once established cells can undergo *temporal variation* in response to changes in the boundary conditions. This can include a change in the sediment budget, as has occurred on many coasts following the postglacial marine transgression, as shelf sediment supplies have become exhausted. It could also be human induced through construction of dams or shoreline structures which interrupt or stop sediment supply. It can also be produced by changes in the driving processes, such as wave climate, wave refraction and climatic conditions, all of which can change both the magnitude and direction of sediment transport within the cell.

Cells can also reach *equilibrium*, when there is essentially zero sediment transport and zero change to the sediment budget over a given time span. This can occur on swash aligned beaches where the wave crest is in equilibrium with the shoreline alignment. On a drift or current aligned shore it would infer zero longshore sediment transport, which is possible but unlikely. On a graded shore the sediment texture is arranged so that it is in equilibrium with the prevailing processes so that no further regrading is required. Swash aligned beaches may take the form of a log spiral (see LOG SPIRAL BEACH) or zeta shape, in locations where waves are refracted around a headland and the shoreline aligns to the spiral of the refracted wave crest. So long as there is no littoral drift they may be considered an equilibrium or closed cell. However on many coasts such systems do release sediment downdrift through pulse boundaries.

Davies (1980) approaches sediment cells from the concept of a sediment store, which has input, throughput and outputs, as well as internal biological supply and loss from attrition. The nature of the store or cell is dependent on the nature, scale and rate of the boundary components and internal dynamics.

Carter (1988) provides the most advanced treatment of what he calls a 'coastal cell', within which is a recognizable compartmentalization of the sediment budget. He goes on to say that this is easy to identify where the cell is restricted to a bay, estuary, river mouth or pocket beach, but many coastal cells have leakages longshore to adjacent systems, onshore to dunes and estuaries or offshore to the inner continental shelf. The cell boundaries may therefore be free or fixed (Table 42). *Fixed* boundaries include impermeable morphological structures such as headlands, shoals, inlets, river mouths. *Free* boundaries are more transparent and therefore less recognizable, as they may result from a change in wave field and direction of transport, rather than a distinctive morphological feature. Both boundary types may either cause the cells to 'divide', 'meet' or 'pulse'. The *divide* boundary occurs at the updrift limit of shoreline erosion, while the *meet* boundary occurs at the downdrift limit of sediment deposition. The *pulse* boundary permits sediment exchange between cells.

Sediment cells can also vary considerably in size and relation to neigbouring cells. Cells may be either independent (e.g. small pocket beach) with fixed boundaries and no leakages, or be nested or cascaded within a series of interconnected (leaking) subcells (e.g. deltaic systems or series of interconnected beaches).

Sediment cells can be identified as a *morphological* unit, particularly when they have fixed boundaries, e.g. an embayed or pocket beach or estuary. They can also be identified based on *sediment texture*, either through the presence of one sediment type, e.g. a coastal dunefield, or

Table 42 Sediment cell definitions and budget

Sediment cell – a closed and balanced sediment budget	
Boundaries	
fixed	morphological structure (e.g. headland, inlet, river mouth, foreland)
free	dynamic divide induced by processes, e.g. change in direction of littoral drift
Boundary types	
divide–	updrift, limit of erosion or sources of sediment
meet–	downdrift, limit of sediment deposition, sink
pulse–	leakage and exchange across boundary
Sediment inputs	
External	
	Terrestrial – river supply, cliff erosion
	Biological – carbonate detritus
	Updrift longshore transport – from neighbouring cell
	Headland bypassing via dunes or subaqueous sand pulses
	Onshore transport from inner shelf, esp. during sea level transgression
Internal	
	Biogenic production
	Chemical production, e.g. ooids and cements
	Beach nourishment
Sediment outputs	
Onshore	barrier – dunes
	– overwashing
	estuary infilling (flood tide deltas)
Longshore	longshore sediment transport
Offshore	inner shelf sand bodies
	submarine canyons
Internal	attrition, solution, cementation (e.g. beachrock)
	beach mining
Sediment balance	
Inputs–outputs = cell erosion, deposition or stable	

Sources: After Komar (1998), Davies (1980) and Carter (1988)

a gradation in the sediment texture resulting from selected longshore transport of size factions. Finally they can be defined by the rates and scale of sediment *transport* within the system, with the boundaries having little or no transport.

References

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

Davies, J.L. (1980) *Geographical Variation in Coastal Development*, 2nd edition, London: Longman.

Komar, P.D. (1998) *Beach Processes and Sedimentation*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.

SEE ALSO: log spiral beach; longshore (littoral) drift; sediment budget

ANDREW D. SHORT

SEDIMENT DELIVERY RATIO

The rate of sediment yield at a specified point in a channel network, expressed as a fraction of the rate of erosion in the contributing catchment, is termed the sediment delivery ratio. Sediment delivery ratios are widely used to adjust SOIL EROSION estimates to account for deposition of sediment as it is transported from its point of origin to, and through, the stream network.

Sediment delivery ratios have been widely used to estimate stream sediment loads from erosion rates predicted by the Universal Soil Loss Equation (USLE) and its successor, the Revised USLE (RUSLE). These empirical equations are designed to predict gross rates of erosion at the soil surface. Because the USLE and RUSLE were

developed from studies of small test plots, they define 'erosion' as the movement of soil particles from one location to another – but, importantly, not necessarily from their point of origin to a stream channel. A fraction of the sediment mobilized by surface erosion will be intercepted (for example, in densely vegetated zones or low-gradient footslopes) before it reaches the channel network. Of the sediment that reaches the channel network, a further fraction will be deposited on the floodplain or stored in the channel. The proportion that is delivered to a sampling point in the channel network – rather than intercepted on the soil surface, deposited on the floodplain, or stored in the channel – is the sediment delivery ratio.

Sediment delivery ratios are commonly estimated from the measured sediment yield (from sediment gauging methods or accumulation in a sediment trap) at a given point in the channel network. This is then divided by the estimated rate of erosion in the surrounding catchment (derived from the USLE/RUSLE or, in some cases, direct field measurements). Thus sediment delivery ratios will not only reflect sediment interception, storage and deposition, but will also reflect any errors made in estimating sediment yields or rates of surface erosion; both are subject to significant uncertainties (Meade 1988; Trimble and Crosson 2000).

Sediment delivery ratios reported in the literature range from over 100 per cent to less than 1 per cent. This variability arises from differences in geomorphic characteristics between catchments, as well as from variations in erosion rates and sediment yields through time at any individual site. Sediment delivery is often highly episodic, and measurements of sediment yield – even when averaged over decades – can be significantly higher or lower than long-term rates of sediment supply to the channel network (e.g. Clapp *et al.* 2000; Kirchner *et al.* 2001).

Nonetheless, systematic relationships have been observed between sediment delivery ratios and catchment morphology and processes. Sediment delivery ratios tend to be higher in catchments where channel slopes and valley sideslopes are steep, and where relief and drainage density are high. Conversely, sediment delivery ratios tend to be lower where sediment sources are far from channels, or are separated from them by sediment-trapping zones (typically characterized by low gradients and dense vegetation). Sediment delivery ratios also tend to be lower where sheet and rill erosion predominate, and higher where gully erosion predominates, because gullies tend to be more directly connected to the channel network.

Sediment delivery ratios also generally decrease as drainage area increases, ranging from roughly 30–100 per cent in 0.1 km^2 catchments to roughly 2–20 per cent in 1,000 km^2 catchments (e.g. Novotny and Olem 1994). This is consistent with the fact that as one moves downstream, channel and valley gradients typically become gentler and floodplains and footslopes typically become wider. All these trends provide greater opportunities for sediment storage, both on hillslopes and in the fluvial system. As one might expect, the lowest sediment delivery ratios are typically observed where rivers emerge from steep mountain fronts and flow out across broad depositional basins.

Sediment yield predictions are often generated by combining USLE or RUSLE estimates of surface erosion with sediment delivery ratios plucked from the literature. This approach is problematic, because sediment delivery ratios vary widely and are not always consistently defined. The denominator of the sediment delivery ratio is sometimes the gross rate of sediment mobilization on the soil surface (as in the USLE/RUSLE); in this case the ratio reflects sediment interception en route to the channel as well as net deposition and storage in the channel network. Alternatively, the denominator is sometimes the rate of sediment supply to the channel network (excluding sediment interception during overland transport, but including sediment production from channel incision or bank erosion); in this case the sediment delivery ratio reflects only the transmission efficiency of the fluvial system. Because sediment delivery ratios may be conceptually defined or operationally measured differently from one study to the next, they should be interpreted with caution.

References

Clapp, E.M., Bierman, P.R., Schick, A.P., Lekach, J., Enzel, Y., and Caffee, M. (2000) Sediment yield exceeds sediment production in arid region drainage basins, *Geology* 28, 995–998.

Kirchner, J.W., Finkel, R.C., Riebe, C.S., Granger, D.E., Clayton, J.L., King, J.G. and Megahan, W.F. (2001) Mountain erosion over 10-year, 10,000-year, and 10,000,000-year timescales, *Geology* 29, 591–594.

Meade, R.H. (1988) Movement and storage of sediment in river systems, in A. Lerman and M. Meybeck (eds) *Physical and Chemical Weathering in Geochemical Cycles*, 165–179, Dordrecht: Kluwer.

Novotny, V. and Olem, H. (1994) *Water Quality: Prevention, Identification, and Management of*

Diffuse Pollution, New York: Van Nostrand Reinhold.
Trimble, S.W. and Crosson, P. (2000) U.S. soil erosion rates – Myth and reality, *Science* 289, 248–250.

JAMES W. KIRCHNER

SEDIMENT LOAD AND YIELD

Normally, sediment discharge is reported in units of mass per unit time, kilograms per second, megagrams per day or megagrams per year. A megagram is equivalent to a tonne, so frequently loads are reported in tons per day or tons per year. Sediment concentration is reported in milligrams per litre or grams per cubic metre of water. The size of particulate material transported by rivers ranges between fine clay and colloidal particles of less than 0.5 micrometres in diameter to large boulders moved during flood events. A distinction is commonly made between bedload and suspended load. This is a distinction based on mode of transport. Bedload consists of the coarser sediment particles that roll, hop or slide along the bed, with more or less continuous contact with the bed, while the suspended load consists of sediment particles that are supported in the flow by the upward components of turbulent currents. The total sediment load is the sum of the bedload and the suspended load. Engineers prefer the distinction between bed-material load and wash load where sediment particles that are finer than those contained in the bed are called wash load and sediment particles that are the same size range as those found in the bed are called bed-material load. The implication is that bed-material load is primarily a function of transporting capacity, boundary shear stress or force available to initiate movement, whereas wash load is commonly a function of variable sources of supply. In either case, there is overlap between the two categories. Church (1983) recommends 1 mm as the appropriate boundary value between bedload and wash load; Walling (1988) recommends 0.2 mm as the cutoff between bedload and suspended load.

Fluvial sediment transported as bedload consists almost entirely of inorganic material, except in steep, forested watersheds where large organic debris is a highly significant component. The inorganic bedload normally resembles the local bedrock in terms of mineral composition except in glacially disturbed watersheds. The finer sediment transported in suspension commonly incorporates a proportion of organic material. The mineralogy of these sediments may bear little resemblance to the local bedrock because of selective detachment and transport processes, chemical weathering processes which disintegrate the rock and secondary redistributed glacigenic sediments. Also the size distribution of the suspended sediment will be considerably enriched in clay-sized particles and organic matter when compared with the source material. Enrichment in fine material has important implications for transport of contaminants which tend to be adsorbed onto finer particles as these finer particles have greater specific surface areas and cation exchange capacities.

Sediment transport capacity load or non-capacity load is a classification of the load of sediment being carried by a stream. If a stream has excess or unsatisfied capacity it is said to be carrying a non-capacity load; if it is carrying as much material as available stream energy permits, it is carrying a capacity load. The enormous capacity for the transport of fine sediment is exemplified by rivers carrying hyperconcentrated flows. Almost all streams carry a non-capacity load of fine sediments. For particles larger than some critical diameter, streams carry a capacity load. A commonly accepted critical particle diameter is around the silt/sand boundary (or about 0.063 mm).

Sediment transport formulae are still evolving, especially bedload formulae. Bedload transport commences when the rate of water flow reaches a magnitude adequate to exert a critical tractive force. Du Boys introduced a formula for movement of bedload of the form

$$Q_b = C_b \cdot \tau(\tau - \tau_c)$$

where Q_b = bedload transport per unit width per unit time, C_b = a special sediment parameter, τ = shear stress, and τ_c = critical shear stress for start of material movement.

Suspended load transport can be considered as an advanced stage of bedload transport by which particles in saltation are caught by the upward component of the turbulent velocity and are kept in suspension. The fundamental equation of sediment suspension by fluid turbulence is

$$c\omega = -\beta \cdot \varepsilon \cdot dc/dy$$

where c = concentration of suspended load in dry weight per unit volume, ω = settling velocity of the particles, β = sediment transfer coefficient, ε = exchange coefficient for suspended load, and γ = vertical distance from the stream bed.

There are problems of application because ε varies through the vertical, and the expression for the variation in suspended sediment concentration with depth is fairly complex. Nevertheless, suspended sediment load can be predicted more accurately than the bedload.

It is also possible to combine bedload and suspended load transport formulae in one expression for total load. The fundamental reference is Einstein (1950) but there are numerous papers still appearing in *Water Resources Research* debating improvements in these formulae.

Milliman and Syvitski (1992) admit that no one knows exactly how much sediment is discharged to the ocean but they suggest that it is probably of the order of 20 billion tonnes per year. This is 50 per cent higher than the estimate by Milliman and Meade (1983) and is accounted for by the fact that the contribution of small mountainous rivers was overlooked in the earlier estimate. On the other hand, it seems probable that, because of the proliferation of dam construction during the second half of the twentieth century, that a smaller total is in order; they also note that an unknown amount is stored in subaerial parts of sinking deltas. The suggestion is that prior to widespread farming and deforestation, the export of sediment was about 10 billion tonnes per year.

Sediment yield is defined as the total mass of particulate material reaching the outlet of a drainage basin per year. Specific sediment yield is the sediment yield per unit area and, making suitable adjustments for sediment specific gravity and density of packing, this figure can be converted into a depth of sediment removed from the whole surface of the basin per year. A unit of measurement known as the Bubnoff (Fischer 1969), which is equivalent to 1 mm of denudation per 1,000 years has been found to be useful for regional comparisons, though it should always be borne in mind that variable thicknesses of sediment are removed from different parts of each basin and the Bubnoff is only an average statistic. Documented values range from 0.09 to over 13,000 Bubnoffs (0.26 to 36,000 $t\,km^{-2}$). The two extreme cases are from Finland and Taiwan respectively. The total sediment yield should include both bedload and suspended load, but only rarely are measurements of bedload available. In such a case, 10 per cent is normally added to the suspended load. In steep mountainous basins, this is clearly an underestimate and in lowland rivers, this is an overestimate.

Where the sediment load of a river is deposited in a lake or reservoir it may be possible to estimate the total yield quite accurately. The sediment yield from a drainage basin is commonly only a small proportion of the gross erosion in the basin. Much of the eroded material is deposited before it reaches the outlet and the ratio of sediment yield to gross erosion is termed the sediment delivery ratio. The magnitude of the sediment yield is a function of climate, topography, sediment availability, lithology, vegetation cover and land use.

The sediment budget is defined as a method of accounting of sources, sinks and redistribution pathways of sediments in a unit region over unit time (Swanson *et al.* 1982). The method provides a bridge between studies of fluvial sediment transport and of the associated deposits in storage in the basin. The transport of sediment out of the basin is then seen as a residual term after deposition and storage of the sediment within the basin. In general, the storage term gets larger with increasing size of basin. Sediment on its way from source to outlet of a basin gets side tracked in a number of ways, not only into lake, floodplain and channel storage, but also into other hillslope locations. It is in this context that the concept of virtual velocity of sediment becomes important (Church 2002). Virtual velocity can be defined as the velocity of sediment through a reservoir and is simply the inverse of residence time per metre of reservoir length. Residence time per metre equals the mapped volume of sediment per metre divided by the bedload discharge rate.

It is estimated for the United States that the amount of sediment delivered to the oceans is only about 10 per cent of the total amount eroded off the uplands and 90 per cent is stored between the erosion sites and the sea. They provide illustrative data on the importance of short-term scale, decade to century scale and longer timescales of storage. The short-term pattern on the lower Mississippi River shows sediment bei
deposited and stored on the river bed at l
flows and being resuspended and flushed
sea at higher flows. The decade to ce
effect is illustrated by the movemen
mining debris through the Sacra
tem in California. The slug of ... of
according to Gilbert 19 ... ks,
mining during 1855–18 ... 0.
to reach the sea and ca ... ation,
bed levels from tem ... on in
that period. A classic stu ... Stream ... ities for

examines the long-term effects of sediment storage resulting from forest clearance and accelerated erosion in the uplands of Coon Creek basin in Wisconsin in the 1850s. During 1853–1975, 80 million tons of sediment were eroded off the uplands and delivered to storage sites whereas only 5 million tons were transported out of the basin. In British Columbia, Church and Slaymaker (1989) estimate that sediment stored in paraglacial fans, valley fill and floodplains will require several tens of thousands of years to be transported to the sea.

References

Church, M. (1983) Concepts of sediment transfer and transfer on the Queen Charlotte Islands, *Fish/Forestry Interaction Program Working Paper* 2/83, Victoria, BC: Ministry of Forests and Ministry of Environment, Lands and Parks.

——(2002) 'Fluvial sediment transfer in cold regions', in K. Hewitt, M.L. Byrne, M. English and G. Young (eds) *Landscapes of Transition*, 93–117, Dordrecht: Kluwer.

Church, M. and Slaymaker, O. (1989) Disequilibrium of Holocene sediment yield in glaciated British Columbia, *Nature* 337, 452–454.

Einstein, H.A. (1950) The bed load function for sediment transportation in open channel flow, *US Department of Agriculture Paper* No. 1,028.

Fischer, A.G. (1969) Geological time-distance rates: the Bubnoff unit, *Geological Society of America Bulletin* 80, 549–552.

Gilbert, G.K. (1917) Hydraulic mining debris in the Sierra Nevada, *US Geological Survey Professional Paper* 105, Washington, DC: US Geological Survey.

Milliman, J.D. and Meade, R.H. (1983) World wide delivery of river sediment to the oceans, *Journal of Geology* 91, 1–21.

Milliman, J.D. and Syvitski, J.P.M. (1992) Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers, *Journal of Geology* 100, 525–544.

Swanson, F.J., Janda, R.J., Dunne, T. and Swanston, D.N. (1982) Sediment budgets and routing in forested drainage basins, *US Forest Service General Technical Report*, PNW-141, Portland, OR: US Department of Agriculture.

Trimble, S.W. (1983) A sediment budget for Coon Creek basin in the Driftless Area, Wisconsin, 1853–1977, *American Journal of Science* 283, 454–474.

Walling, D.E. (1988) Erosion and sediment yield research: some recent perspectives, *Journal of Hydrology* 100, 113–141.

Further reading

Dietrich, W.E. and Dunne, T. (1978) Sediment budget for a small catchment in mountainous terrain, *Zeitschrift für Geomorphologie Supplementband* 29, 191–206.

Hallet, B., Hunter, L. and Bogen, J. (1996) Rates of erosion and sediment evacuation by glaciers: a review of field data and their implications, *Global and Planetary Change* 12, 213–235.

Hinderer, M. (2001) Late Quaternary denudation of the Alps, valley and lake fillings and modern river loads, *Geodinamica Acta* 14, 231–263.

Hovius, N., Stark, C.P. and Allen, P.A. (1997) Sediment flux from a mountain belt derived by landslide mapping, *Geology* 25, 231–234.

SEE ALSO: sediment rating curve; sediment routing

OLAV SLAYMAKER

SEDIMENT RATING CURVE

A sediment rating curve represents the relation between suspended (see SUSPENDED LOAD) sediment concentration (or discharge) and water discharge at a stream measurement station. It is used for estimating suspended sediment discharge averaged over a period of flow record. The relation may concern 'instantaneous' values or may relate average values over (e.g. daily) time intervals.

In theory, suspended sediment concentration should increase with water discharge because the associated increase in turbulence increases the capacity of the river to carry suspended sediment. In practice, the concentration of suspended sediment, particularly the silt-clay fractions, tends to be more influenced by the sediment supply to the channel from hillslope and riparian erosion processes, which can be 'patchy' in space and time. During runoff events, the relation can vary due to sediment exhaustion and also to phase lags between the sediment and water peaks passing the measurement station. The relation can also vary due to seasonal controls on the supplies of sediment and water, to changes in basin land use, or following extreme hydrological events.

Recognizing that there is no unique relation between suspended sediment concentration and water discharge, a sediment rating curve thus aims to model the conditional mean sediment concentration (as a function of water discharge). This is estimated by sampling a series of concurrent measurements of water discharge, Q, and discharge-weighted sediment concentration, C. Ideally, these samples should be collected over a wide range of water discharge, at rising and falling stages, over all seasons, and over many years so that all the factors inducing variance in the relation are represented in an unbiased fashion.

The traditional approach to fitting a rating curve has been to plot C against Q on log-log graphs. These plots accommodate the large ranges of Q and C observed in rivers, the data-scatter tends to be homoscedastic (i.e. independent of discharge), and the underlying relation often shows a simple power form $C = aQ^b$ (a and b are empirical coefficients) which is linear on a log-log plot and easily modelled with linear regression methods. Note that by using log data, the regression procedure models the geometric conditional mean, rather than the desired arithmetic conditional mean, thus a correction factor is required when transforming the rating curve back from log values (Ferguson 1986).

The power law approach should be applied with caution (e.g. Walling and Webb 1988), since large errors can arise because the least-squares curve, while appearing to fit the overall dataset well, may fit poorly at high discharges, and it is these discharges that usually transport the bulk of the sediment load. In such cases, other curve-fitting techniques such as non-linear regression, Locally Weighted Scatterplot Smoothing (LOWESS), or simply segmented curves based on subsets of the data perform better.

References

Ferguson, R.I. (1986) River loads underestimated by rating curves, *Water Resources Research* 22, 74–76.

Walling, D.E. and Webb, B.W. (1988) The reliability of rating curve estimates of suspended sediment yield: some further comments, *International Association of Hydrological Sciences Publication* 174, 337–350.

D. MURRAY HICKS

SEDIMENT ROUTING

The process through which sediment is transported downstream following a specific path or route. The sediment is fluvial sediment, which includes both bedload and suspended load. The path or route, of the sediment may be the course of the natural channel, an artificial canal or a restored channel. Sediment routing is closely related to sediment transport. Where sediment transport may be concerned with the details of sediment movement, sediment routing defines the path of that sediment on a channel reach or watershed scale.

Sediment routing models are used in measuring the SEDIMENT BUDGET for a watershed. The routing model quantifies sediment sources and sinks in the watershed, and how much and how fast that sediment is transferred downstream by river reaches. One of the first steps in measuring the sediment budget in a watershed is the creation of a map of the paths travelled by the sediment.

Sediment routing is affected by landscape changes. Where roads are built in forests, there is an immediate effect on the route travelled by the fluvial sediments. Where roads cut across channels, the path of the water and sediment often changes to follow the road instead of continuing down the natural channel. The result may be either deposition on the road surface or increased road surface erosion. Either way, the path of the channel has been changed, altering the sediment routing processes in the watershed.

Artificial canals are built to route sediment through a city without causing damage. Trapezoidal concrete channels route both water and sediment through residential areas without causing flooding or other damage during high flow events. These types of canals are most common where towns are situated next to mountains, for example, Los Angeles, California and Albuquerque, New Mexico. When a large flow event occurs, the canals are filled with material flowing off the mountain side.

Sediment routing is an important consideration when planning a river restoration project, and restorations are often undertaken when the transport of sediment through the channel has been determined to be too slow or too fast, resulting in either erosion or sedimentation of the channel bed. Natural channels are altered to change the sediment route through the addition of sediment sinks and bedforms. Often, both the path and slope of the channel may be altered in an attempt to change the rates of sediment and water transport. In the case of the Florida Everglades, a meandering channel was straightened. The route that the sediment travelled went from a winding path to a straight line. Subsequent effects to the ecosystem were disastrous, and current work is attempting to re-create the original route.

Further reading

Benda, L. and Dunne, T. (1997) Stochastic forcing of sediment routing and storage in channel networks, *Water Resources Research* 33(N12), 2,865–2,880.

Madej, M.A. (1993) Development, implementation, and evaluation of watershed rehabilitation in Redwood National Park, in *Watershed and Stream Restoration Workshop: Shared Responsibilities for*

Shared Watershed Resources, Symposium Proceedings of the American Fisheries Society, 75–82, Portland, OR.

SEE ALSO: sediment budget

JOANNA C. CURRAN

SEDIMENT WAVE

A sediment wave is a transient zone of sediment accumulation in a river channel that is created by sediment input and does not originate solely from variations in channel topography. Similar terms are 'sediment slug' or 'pulse'. Sediment waves have a minimum spatial scale measured in channel widths and a minimum volumetric scale corresponding to major bars (see BAR, RIVER); they exist over a number of hydrographic events (Nicholas *et al.* 1995; Lisle *et al.* 2001). A sediment wave is not necessarily a body of bed material moving *en masse* downstream, but evolves as a disturbance in the interactions between flow, channel topography and the transport of bed material that can be contributed from any part of the basin, including the input immediately responsible for forming the wave.

Sediment waves evolve by various degrees of dispersion and translation, depending on the form and sedimentology of the channel. Dispersion is the spreading of the wave as its apex lowers and remains stationary and its trailing edge remains stationary or is accreted upstream. Translation is the downstream advancement of the wave, including its apex and trailing edge. A bed material wave that evolves only by dispersion is still regarded as a wave. Dispersion dominates the evolution of bed material waves in quasi-uniform, gravel-bed channels (see GRAVEL-BED RIVER), where the Froude number at significant sediment-transporting flows is generally high (<1); translation of waves can be important in sand-bed channels where the Froude number is low ($\ll 1$) (Lisle *et al.* 2001). Well-documented examples of these contrasting behaviours are provided by Sutherland *et al.* (2003) and Meade (1985). Wave material that is finer than ambient bed material can promote wave translation, but sand waves in steep, gravel-bed channels tend to evolve primarily by dispersion (Lisle *et al.* 2001).

The foregoing applies to quasi-uniform channels, but other features and processes in natural channels may promote wave translation. DEBRIS FLOWS, for example, translate sediment downstream during single events. The downstream spread of a sediment wave through a series of sedimentation zones (Church 1983), where deposition is locally enhanced by valley-scale topography, could manifest wave translation. Advancing zones of increased transport at the leading edge of a wave may activate sediment stored in unstable reaches and propagate a more pronounced wave downstream (Wathen and Hoey 1998).

The relative dominance of dispersion and translation is important for river resources. Dispersion of sediment inputs attenuates but prolongs sediment impacts downstream, whereas translation propagates pronounced sequences of impact and recovery.

References

Church, M. (1983) Pattern of instability in a wandering gravel bed channel, in J.D. Collinson and J. Lewin (ed.) *Modern and Ancient Fluvial Systems*, 169–180, Oxford: Blackwell Scientific.

Lisle, T.E., Cui, Y., Parker, G., Pizzuto, J.E. and Dodd, A.M. (2001) The dominance of dispersion in the evolution of bed material waves in gravel bed rivers, *Earth Surface Processes and Landforms* 26(13), 1,409–1,420.

Meade, R.H. (1985) Wavelike movement of bedload sediment, East Fork River, Wyoming, *Environmental Geology Water Science* 7(4), 215–225.

Nicholas, A.P., Ashworth, P.J., Kirkby, M.J., Macklin, M.G. and Murray, T. (1995) Sediment slugs: large-scale fluctuations in fluvial sediment transport rates and storage volumes, *Progress in Physical Geography* 19(4), 500–519.

Sutherland, D.G., Hansler, M.E., Hilton, S. and Lisle, T.E. (2003) Evolution of a landslide-induced sediment wave in the Navarro River, California, *Geological Society of America Bulletin*, 114, 1036–1048.

Wathen, S.J. and Hoey, T.B. (1998) Morphologic controls on the downstream passage of a sediment wave in a gravel-bed stream, *Earth Surface Processes and Landforms*, 23, 715–730.

SEE ALSO: sediment routing

THOMAS E. LISLE

SEDIMENTATION

The term sedimentation refers to the settling of solids from suspension in a fluid. In geomorphology the fluid is typically water (fluvial, lacustrine or marine sedimentation) or air (aeolian sedimentation). The fundamental process of sedimentation is described by Stokes' law which describes the settling of spherical particles from a still fluid.

This is expressed as $V = (2gr^2)(d_1 - d_2)/9\eta$ where: V is the particle fall velocity (cm s^{-1}), g is acceleration due to gravity (cm sec^{-2}), r is the radius of the particle (cm), d_1 is the density of the particle (g cm^{-3}), d_2 is the density of the fluid (g cm^{-3}) and η is the viscosity of the fluid (dyne sec cm^{-2}). In natural systems this simple relation is complicated by the non-spherical nature of most sedimentary particles, and by the fact that particles are typically deposited from a moving fluid column. In a moving fluid the vertical component of turbulent eddies transfers momentum to the particle that may exceed the velocity of gravitational settling so that the particle remains suspended. As flow velocities decline successively finer particles settle out of the column so that theoretically the typical sedimentary structure associated with sediments deposited from a waning flow is a normally graded bed which fines upwards. In practice, the nature of the source sediment, temporal and spatial variability in flow, and post depositional modification, significantly complicate the nature and interpretation of deposits. For a good review of controls on sedimentation and the interpretation of deposits see Leeder (1992).

As a consequence of the difficulties of developing physical models of sediment transport and deposition, geomorphologists have developed empirically based generalizations to describe sedimentation in particular systems. Examples include the Hjulström curve (Hjulström 1935) which predicts entrainment and deposition threshold velocities for sediments of varying size in fluvial systems. More work has focused on defining conditions for the initial entrainment of sediment than subsequent sedimentation. (See for example Shields (1936) diagram or work by Bagnold (1941) on aeolian entrainment.) Threshold velocities for deposition are lower than those for entrainment due to the role of inertia and bed packing in limiting initial motion. For example bedload sediments in Turkey Brook, England show depositional thresholds that are only 35 per cent of entrainment thresholds (Reid and Frostick 1994).

Geomorphological studies of sedimentation can be divided into those which aim to analyse and model contemporary sedimentary processes, and those which draw on this understanding to interpret past environments and processes from sedimentary evidence. The former, in the tradition of the work by Shields and Bagnold, constitute a central part of modern process geomorphology. The insights of the latter approach, often classified as Quaternary science or historical geomorphology, are equally necessary for the proper understanding of contemporary landscapes.

Within geomorphology a broader definition of sedimentation is generally accepted which includes the deposition of sediments from beneath glaciers and ice sheets (glacial sedimentation) and deposition by mass movement processes on slopes. Sedimentation in glacial environments is complex, ranging from purely glacial deposition of sediment, where there is a close link between style of deposition and the dynamics of the glacial ice, through waterlain tills deposited below floating ice, to extensive fluvial and aeolian sedimentation of glacially derived material. A good review of glacial sedimentation is given by Hambrey (1994).

In the broader sense referred to above, sedimentation is taken to be synonymous with deposition of sediment and is consequently central to the study of a wide range of depositional landforms and environments. Sedimentation is the process which defines the end point of the sequence of sediment EROSION, sediment transport and sediment deposition. As such, accumulation of sediment in sedimentation zones provides an integration of upstream/upglacier/upwind geomorphic action. Where system stability allows that the locus of sedimentation remains constant over time the resulting depositional sequence provides a valuable stratigraphic archive of changing rates of sediment flux. These records are of particular importance in determining process rates over geomorphologically significant timescales, extending beyond the few years of most monitored datasets. Lake sediments are a good example of a stable locus of sedimentation and numerous studies of lake sedimentation have inferred long-term catchment sediment yields from lake sediment volumes and chronological control on stratigraphy (e.g. Desloges and Gilbert 1998). In appropriate contexts lake sediments can be used not only to measure sediment flux, but also to identify changes in style of sedimentation which indicate change in the balance of sediment delivery processes. For example, Menounos (2000) used distinct coarse layers in the sediments of a high alpine lake in Colorado to reconstruct the frequency of debris flow activity in the catchment.

Church and Jones (1982) used the term sedimentation zones to describe river reaches where sediment characteristically accumulates, which are separated by reaches where sediment transport dominates. Sedimentation zones are large-scale stores of sediment within the landscape system. The concept can be to some extent generalized. Sedimentation is not uniform across landscape systems but is typically concentrated in zones where the landscape morphology is such that sedimentation is favoured. In the strict definition of sedimentation these areas are those where lower fluid velocities or physical retention of sediment promote sedimentation, for example lakes, or areas of preferential aeolian sedimentation around major topographic obstacles. A good example of the latter are the sand dune systems of Great Sand Dune National Monument in Colorado where sands deflated from alluvial surfaces accumulate against a downwind mountain front (Plate 120). Taking the broader definition of sedimentation, breaks of slope, which promote the accumulation of slope deposits, may be identified as areas of preferential sedimentation. In alpine environments for example cirque basins can be identified as morphological units characterized by local sedimentation.

One common usage in the literature is accelerated sedimentation referring to increased rates of sedimentation usually as a consequence of human action. Accelerated sedimentation is the necessary consequence of enhanced erosion upstream due to human modification of land use, or of increased deposition associated with human modification of the fluid flow. The most common impacts include mining and intensification of upland agriculture. The classic study of mining impacts was carried out by Gilbert (1917) on the impacts of hydraulic gold mining in the Sierra Nevada. Between 1853 and 1884 over 1 billion

Plate 120 Alluvial sediments in the foreground provide the sediment source for aeolian sedimentation of the dune systems against the mountain front, Great Sand Dune National Monument, Colorado

m^3 of sediment was produced by mining leading to major fluvial aggradation of upland valleys. This sediment continued to be reworked through the Sacramento valley throughout the twentieth century (James 1999). Xu (1998) records a 25 fold increase in sedimentation rates in the Lower Yellow River in China spanning the past 2,200 years associated with a 40 fold increase in population upstream and consequent intensification of agriculture and engineering works on the river. Similarly, O'Hara *et al.* (1993) suggested that significant increases in lake sedimentation rates in a Mexican highland lake were due to over-intensification of pre-Hispanic agricultural practices.

Phases of accelerated sedimentation are not confined to fluvial systems. Severe desertification in north-central China is occurring as a result of human modification of land use leading to accelerated aeolian sedimentation due to encroachment of dune systems (Fullen and Mitchell 1994).

In general the most dramatic increases in sedimentation rates across a range of different environments, are typically seen within the past 200 years, associated with mechanization and increased rate of land use change in response to rapid population growth.

A related term often used in this context, particularly with reference to infill of river channels and reservoirs, is siltation. The strict definition of siltation is the sedimentation of silt-sized particles but the term is also used more generally to refer to the infill of channels and basins with fine-grained sediment. In areas of rapid erosion reservoir siltation can significantly reduce reservoir capacity and represents a significant economic cost (Palmieri *et al.* 2001) necessitating erosion control measures in the catchment. Fine-grained sedimentation causes particular problems in upper reaches of salmon streams (Hartman *et al.* 1996) as the sedimentation clogs the gravel interstices and inhibits spawning of the returning fish. In areas of the Pacific North West of North America where commercial logging coexists with economically important salmon runs the fine-grained sedimentation associated with logging-related slope failures is a major source of land use conflict.

Long-term changes in sedimentation rate also occur naturally without direct human impact. One of the main drivers is climate change, either directly through impact on weathering and erosion rates, or indirectly through climate driven vegetation change (e.g. Evans 1997; Xu 1998). One major shift in sedimentation rates characteristic of

formerly glaciated areas is a peak in sedimentation associated with deglaciation. PARAGLACIAL sedimentation, conditioned by the former presence of glacial ice, commonly occurs at rates of at least an order of magnitude above subsequent equilibrium rates of sedimentation (Hinderer 2001).

Increases in sedimentation rate at a point have two fundamental causes. One is changes in the nature of erosion so that the sediment load is increased and rates of sedimentation increase without necessary changes in the style or location of sedimentation. The second main cause is associated with anthropogenic changes to the nature of the landscape system which change the balance of sedimentation and sediment transport for a particular location. An understanding of sedimentation and sedimentary processes are important to the geomorphologist in many ways. There are many practical applications in controlling and mitigating anthropogenically induced sedimentation, and the sediments are valuable archives of past rates and styles of sedimentation. Fundamentally, however, sedimentation is necessarily linked to erosion (the source of the sediment) and it is the balance of erosion and sedimentation across space and over long timescales which determines the geomorphology of contemporary landscapes.

References

Bagnold, R.A. (1941) *The Physics of Blown Sand and Desert Dunes*, London: Methuen.
Church, M. and Jones, D. (1982) Channel bars in gravel-bed rivers, in R.D. Hey, J.C. Bathurst and C.R. Thorne (eds) *Gravel-bed Rivers*, Chichester: Wiley.
Desloges, J.R. and Gilbert, R. (1998) Sedimentation in Chilko Lake: a record of the geomorphic environment of the eastern Coast Mountains of British Columbia, Canada, *Geomorphology* 25, 75–91.
Evans, M. (1997) Temporal and spatial representativeness of alpine sediment yields: Cascade mountains, British Columbia, *Earth Surface Processes and Landforms* 22, 287–295.
Fullen, M.A. and Mitchell, D.J. (1994) Desertification and reclamation in North-Central China, *Ambio* 23(2), 131–135.
Gilbert, G.K. (1917) Hydraulic mining debris in the Sierra Nevada, *US Geological Survey Professional Paper* 105, Washington, DC: US Geological Survey.
Hambrey, M. (1994) *Glacial Environments*, Vancouver: UBC Press.
Hartman, G.F., Scrivener, J.C. and Miles, M.J. (1996) Impacts of logging in Carnation Creek, a high energy coastal stream in British Columbia, and their implication for restoring fish habitat, *Canadian Journal of Fisheries and Aquatic Sciences* 53, 237–251.
Hinderer, M. (2001) Late Quaternary denudation of the Alps, valley and lake fillings and modern river loads, *Geodinamica Acta* 14, 231–263.
Hjulström, F. (1935) Studies of the morphological activity of rivers as illustrated by the river Fyris, *Bulletin of the Geological Institute, University of Uppsala* 25, 221–257.
James, A. (1999) Time and the persistence of alluvium: river engineering, fluvial geomorphology and mining sediment in California, *Geomorphology* 31, 265–290.
Leeder, M.R. (1992) *Sedimentology: Process and Product*, London: Chapman and Hall.
Menounos, B. (2000) A Holocene debris-flow chronology for an alpine catchment, Colorado Front Range, in O. Slaymaker (ed.) *Geomorphology, Human Activity and Global Environmental Change*, 117–149, Chichester: Wiley.
O'Hara, S.L., Street-Perrott, F.A. and Burt, T.P. (1993) Accelerated soil erosion around a Mexican highland lake caused by pre-Hispanic agriculture, *Nature* 362, 48–51.
Palmieri, A., Shah, F. and Dinar, A. (2001) Economics of reservoir sedimentation and sustainable management of dams, *Journal of Environmental Management* 61, 149–163.
Reid, I. and Frostick, L.E. (1994) Fluvial sediment transport and deposition, in K. Pye (ed.) *Sediment Transport and Depositional Processes*, 89–144, Oxford: Blackwell.
Shields, A. (1936) Anwendung der Aehnlichkeitsmechanik und der Turbulenzforschung auf die Geschiebebewegung, *Mitteilung der Preussischen Versuchsanstalt für Wasserban und Schiffban*, Heft 26, Berlin.
Xu, J. (1998) Naturally and anthropogenically accelerated sedimentation in the lower Yellow River, China over the past 13000 years, *Geografiska Annaler* 80A(1), 67–78.

SEE ALSO: alluvial fan; bar, river; dune, aeolian; erosion; glacial deposition; paraglacial

MARTIN G. EVANS

SEISMOTECTONIC GEOMORPHOLOGY

Seismotectonic geomorphology is the study of landforms produced by earthquakes. It combines the results of seismotectonic, geomorphic and palaeoseismological research. Palaeoseismology deals with the age, frequency and size of prehistoric earthquakes (Wallace 1981). The palaeoseismic record includes strong (M >6.5) and very strong (M >7.8) earthquakes, since geological effects of moderate or weak earthquakes are rarely preserved in the near-surface zone. Seismic activity is associated with active faulting. Faults are considered active when they show potential or probability of future displacements in the present tectonic setting, or may have displacement within a future period of concern to humans, i.e. they have ruptured during the Holocene (active faults) or the Quaternary (potentially active faults). Active faults are usually segmented, each segment showing a different history of movement. Large earthquakes of a characteristic size repeatedly rupture the same part or segment of a fault.

Earthquakes producing recognizable surface deformation are called morphogenic earthquakes, whereas deposits or landforms formed during an earthquake are described as coseismic, as opposed to delayed-response features that follow the seismic event. Geomorphic features occurring both on-fault and off-fault form either primary (resulting from coseismic slip) or secondary (produced by earthquake shaking, like rockfalls or deformed tree rings) evidence of seismicity (McCalpin 1996). Sediments deformed during seismic shaking are called seismites.

Evidence of present and past earthquakes include deformation of the ground surface along seismogenic faults (fault scarps, fissures, sag ponds, offset stream valleys, shutter ridges, folded terraces, deformed alluvial fans, river reversals, fractured cave structures, displaced beach ridges, coral platfroms, delta plains or wave-cut notches), large-scale features of sudden uplift or subsidence above plate-boundary faults (warped river terraces, elevated shorelines, drowned tidal marshes, emerged or subsided coral reefs), as well as stratigraphic or geomorphic effects of ground shaking or tsunamis far from the seismogenic fault (i.e. landslides, slumps, rockfalls, liquefaction features like mud volcanoes or sand-blow deposits).

The primary geomorphic evidence of seismic activity associated with *normal faulting* are fault scarps (see FAULT AND FAULT SCARP). These scarps range in size from mountain fronts up to 1 km high, cut on bedrock, to centimetre-scale scarplets in unconsolidated sediments. Simple (single-event) fault scarps are formed almost instantaneously, and attain heights from a decimetre to a few metres per event. In normal or reverse faulting such scarps face in the direction of slip, whereas during strike-slip faulting they face in different directions. At the base of recent fault scarps, closed basins or sag ponds may develop. Some scarplets, called earthquake rents, reverse scarplets or cicatrices, parallel the base of the scarp but face uphill. Horsts and grabens, as well as unpaired normal faults creating half-grabens (fault-angle depressions) are also very common. Coseismic displacement on normal faults is characterized by greater subsidence of the hanging wall as compared to the size of uplift of the footwall. Scarp degradation is affected by both lithologic and (micro)climatic factors. In semi-arid climate, the free face in unconsolidated sediments becomes completely destroyed in a timespan ranging from one day (1983 Borah Peak earthquake, Ms = 7.3) to one to two thousand years (Crone *et al.* 1987; Wallace 1977). Fractured bedrock scarps will gradually degrade to an angle of repose that can be maintained for one million years or so. Scarps formed by more than one earthquake are called compound, composite or multiple-event scarps. During each earthquake, a colluvial wedge is shed from the scarp, being subsequently covered in interseismic periods by soils that can be dated by different techniques. The fault-induced incision into the upthrown block after faulting can create tectonic terraces that diverge downstream and abruptly terminate against the fault scarp.

Thrust earthquakes occur frequently in fold-and-thrust belts (e.g. Algeria, 1980), at convergent continental plate boundaries (Iran, 1978; Armenia, 1988) or in regions near transpressive bends of strike-slip faults. During thrusting, hanging wall uplift usually exceeds footwall subsidence, and the area affected by coseismic deformation depends on the size of displacement, fault geometry and the rigidity of the deformed crust. The size of displacement diminishes towards the thrust tip, hence, the hanging wall is usually folded adjacent to that tip. Many M < 7 earthquakes may not be accompanied by any geomorphic expression. Typical landforms produced by thrusting are thrust fault scarps which are usually more sinuous and irregular as compared to other fault types, and are composed of short, disconnected segments, or produce a continuous but zigzag-like trace on the scale of metres. These scarps show varied morphology due to mixed low-angle faulting and folding. Seven to eight types of thrust fault scarps have been distinguished. Steeply dipping reverse faults in bedrock form simple thrust scarps, whereas those in unconsolidated sediments result in hanging wall collapse scarps. Low-angle thrust faults produce pressure ridges of shape depending on surface material rheology and the magnitude of slip. In more cohesive materials, pressure ridges have smoother fronts and may display backthrusts or represent low-angle pressure ridges, but as thrust displacement decreases below 1 m, all pressure ridge types merge into a single type of a small moletrack. An increasing oblique component of slip produces *en echelon* pressure ridges or oblique tension gashes in pressure ridge front. Surface thrusting deforming flat terrains is usually expressed as a wide gentle warp of fluvial/marine terrace surfaces. Thrust faulting in bedrock produces an overhanging scarp, but in unconsolidated deposits such overhangs collapse very quickly, creating a free face and a steep debris slope. Numerous degraded reverse fault scarps are asymmetric, with the steepest part of the scarp lying downslope of the scarp midpoint, whereas the normal fault scarps typically show symmetric cross profiles. High escarpments formed by repeated thrusting are usually obscured by a long chain of landslides. Thrust faults are difficult to recognize where cutting high relief and irregular topography, whereas broadly distributed thrusting, surface warping and folding cannot be detected unless planar geomorphic features are deformed.

Active folding is a coseismic process related to faulting on a blind thrust or other dip-slip fault at depth. It can also induce seismicity due to flexural slip during folding. Geomorphic manifestations of surface folds are deformed fluvial channels and terraces. Hanging wall ramp folds can generate distinct facets at the top of scarps, resembling those of normal fault scarps. These facets, however, may be independent of the palaeoseismic history of the fault. Some propagation folds formed at thrust tips may also produce multifaceted scarp profiles.

Major seismogenic *strike-slip faults* are associated either with plate boundaries or are located within intraplate settings, at the boundaries of continental microplates. Landforms made by (palaeo)earthquakes along active strike-slip faults include: linear valleys (they can be created by simple deflection of streams along the fault trace, even without brecciation), offset, beheaded or deflected valleys and streams, offset ridges, sags and sag ponds (related to downwarping between two strands of the fault zone), shutter ridges (formed where a fault displaces ridge crests on one side of the fault against gullies on the other side; usually occurring where a fault breaks through the pre-existing pressure ridges), pressure ridges (small warped areas formed by compression between multiple traces in a fault zone), topographic benches (elevated, flat, gently warped or tilted areas, usually formed due to the displacement between several fault segments or splays in the fault zone), fault scarps of minor to moderate height, and small-scale horsts, grabens and pull-apart basins. The fault trace is frequently composed of a wide zone of alternating tension gashes (extensional) and moletracks (compressional) that strike obliquely to the general fault strike. Landforms typically used to estimate palaeoseismic offset are: fluvial terraces, stream channels and alluvial fans. A minimum slip rate, e.g. of the San Andreas fault, California, has been estimated based on offset alluvial fans at 10–35 mm/yr (Keller *et al.* 1982). The maximum single-event stream offset of 9.5 m was produced by the 1857 Ft. Tejon, San Andreas fault, earthquake of M = 8 (McCalpin 1996).

References

Crone, A.J., Machette, M.N., Bonilla, M.G., Lienkaemper, J.J., Pierce, K.L., Scott, W.E. and Bucknam, R.C. (1987) Surface faulting accompanying the Borah Peak earthquake and segmentation of the Lost River Fault, Central Idaho, *Bulletin Seismological Society of America* 77, 739–770.

Keller, E.A., Bonkowski, M.S., Korsch, R.J. and Shlemon, R.J. (1982) Tectonic geomorphology of the San Andreas fault zone in the southern Indio Hills, Coachella Valley, California, *Geological Society of America Bulletin* 93, 46–56.

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

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

——(1981) Active faults, paleoseismology, and earthquake hazards in the western United States, in D.W. Simpson and P.G. Richards (eds) *Earthquake Prediction – An International Review*, 209–216, Washington, DC: American Geophysical Union.

Further reading

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

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

Schumm, S.A., Dumont, J.F. and Holbrook, J.M. (2000) *Active Tectonics and Alluvial Rivers*, Cambridge: Cambridge University Press.

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

Wallace, R.E. (ed.) (1990) *The San Andreas Fault System, California*, Washington, DC: US Geological Survey Professional Paper 1,515.

SEE ALSO: crustal deformation; fault and fault scarp

WITOLD ZUCHIEWICZ

SELF-ORGANIZED CRITICALITY

Self-organized criticality is an approach to understanding non-linear systems initiated by Per Bak and colleagues, as explained in his book *How Nature Works* (Bak 1997). Self-organized criticality is one of a whole host of linked new approaches, including CHAOS THEORY, complexity theory (see COMPLEXITY IN GEOMORPHOLOGY) and FRACTALS, which aim to provide better explanations of the complex behaviour of non-linear natural systems.

Self-organized criticality is used to explain the behaviour of many complex natural systems which seem to evolve into a poised or critical state, far from equilibrium. Per Bak uses the helpful analogy of a self-organized system being like a sandpile created by dropping grains of sand onto a flat surface. During the initial stages of development of the sandpile, predicting the behaviour of the pile is relatively simple and depends upon the physical properties of the individual grains. As the pile grows, however, avalanches start to occur, whose behaviour is complex and cannot be predicted by the characteristics of the individual grains. At this point the system becomes a complex, self-critical one – whose behaviour can only be understood by considering the properties of the whole pile (a holistic approach) rather than from a reductionist description of the behaviour of individual grains. There are, however, many concepts of self-organization used in science which have subtly different meanings and which are based on different interpretations of natural systems (as elucidated for geomorphology by Phillips 1999). Geomorphologists have used a range of these concepts.

In recent years geomorphologists have had a great interest in ideas such as self-organized criticality which may help explain many of the complex landscapes we see around us. Why and how, for example, do regular patterns such as river networks, rills, stone polygons, beach cusps and dune systems develop? Reductionist approaches to these questions have hoped that studying the basic physics of processes operating at the microscale could provide a general answer. However, such approaches have not often proved able to successfully link process and pattern across different scales. Could such patterns instead be seen to be examples of self-organized criticality, where regular patterns have emerged out of the complex behaviour of smaller scale processes? Many geomorphologists have used cellular models to investigate such systems, in which simple rules are applied to describe the interactions of neighbouring cells. As the models are run, patterns at a larger scale emerge from these simple rules.

Several examples illustrate the use of cellular models to investigate self-organized criticality in geomorphic systems. Werner and Fink (1993) investigated the formation of beach cusps, finding that they could be simulated from a cellular model based on the interaction of water flow, sediment transport and morphological change. Similarly, and at a larger scale, Rodriguez-Iturbe and Rinaldo (1997) use models for flow and sediment transport to develop fluvial networks. The resultant networks have fractal and multifractal properties and are seen by Rodriguez-Iturbe and Rinaldo to be the product of self-organizing processes. In a recent study, de Boer (2001) has built a cellular model to simulate the long-term evolution of a fluvial landscape. Using simple rules to model sediment transport between adjacent cells, a record of total sediment yield is created which has complex magnitude and frequency properties which cannot be predicted from the simple, local rules. Thus the sediment dynamics of the modelled landscape are an emergent property of the entire system resulting from the local interaction of individual cells. For geomorphologists, the insight that complex behaviour may be the result of internal interactions rather than external forcings is highly important, and complicates our search for the geomorphic imprint of external forcings such as climate change.

However, despite these promising model simulations it is as yet unclear whether self-organized criticality really exists in natural geomorphic systems. The behaviour of modelled systems cannot simply be applied to natural systems, for which we do not always have enough data and in which external forcings may also play a role. Werner (1999) suggests that hierarchical models might be better suited to modelling complex natural landform patterns which self-organize in temporal hierarchies.

References

Bak, P. (1997) *How Nature Works: The Science of Self-organized Criticality*, Oxford: Oxford University Press.
de Boer, D.H. (2001) Self-organisation in fluvial landscapes: sediment dynamics as an emergent property, *Computers and Geosciences* 27, 995–1,003.
Phillips, J.D. (1999) Divergence, convergence and self-organization in landscapes, *Annals, Association of American Geographers* 89, 466–488.
Rodriguez-Iturbe, I. and Rinaldo, A. (1997) *Fractal River Basins: Chance and Self-organization*, Cambridge: Cambridge University Press.
Werner, B.T. (1999) Complexity in natural landform patterns, *Science* 284, 102–104.
Werner, B.T. and Fink, T.M. (1993) Beach cusps as self-organised patterns, *Science* 260, 968–971.

Further reading

Favis-Mortlock, D. (1998) A self-organizing dynamic systems approach to the simulation of rill initiation and development of hillslopes, *Computers and Geosciences* 24, 353–372.
Hallet, B. (1990) Self-organisation in freezing soils: from microscopic ice lenses to patterned ground, *Canadian Journal of Physics*, 68, 842–852.
Kauffman, S. (1995) *At Home in the Universe: The Search for Laws of Self-organisation and Complexity*, Oxford: Oxford University Press.
Werner, B.T. and Hallet, B. (1993) Numerical simulation of self-organized stone stripes, *Nature* 361, 142–145.

HEATHER A. VILES

SENSITIVE CLAY

The basic idea of sensitivity is that the structure of the clay soil system has an effect on the properties, so that once the structure is destroyed (by remoulding) a new set of properties is observed. This is probably true in just about all clays although in heavily overconsolidated (see OVERCONSOLIDATED CLAY) systems the effect will be negligible; but in the QUICKCLAYS structure is all-important and when it is destroyed most of the strength properties disappear. The high sensitivities are of interest in applied geomorphology because of associated ground failures and landslides.

The sensitivity of a clay is the ratio of the undisturbed strength to the remoulded strength. It appears to have been first defined by Karl Terzaghi in 1944. The next development is due to Skempton and Northey (1952) and it is this paper which launched the scientific study of sensitivity. There had been Scandinavian experiences with sensitive clays for many years but the Skempton and Northey paper marks a critical beginning. They divided the sensitivity range:

About 1,	insensitive clays
1–2,	low sensitivity
2–4,	medium
4–8,	sensitive
>8,	extra-sensitive
>16,	quickclays

Much of the discussion about sensitivity in clays concerns the mechanism by which it arises. There has been considerable discussion on this topic and many factors influencing sensitivity have been proposed (see Mitchell and Houston 1969); the factors have been nicely ordered by Quigley (1979).

Factors affecting sensitivity

Factors producing high undisturbed strength and high sensitivity:

1 Depositional flocculation, including low electro-kinetic potential, high sediment concentration, divalent cation adsorption.
2 Slow increase in sediment load.
3 Cementation bonds, including carbonates and amorphous sesquioxides

Factors producing low remoulded strength and high sensitivity:

1 High water content (greater than the liquid limit), little consolidation.
2 Low specific surface of soil grains, high silt content or high rock flour content in the clay fraction. High primary mineral, low clay mineral content.
3 High electro-kinetic (zeta) potential, low salinity via leaching, organic dispersants, inorganic dispersants, high monovalent cation adsorption relative to divalent cations.
4 Low amorphous content.
5 Low smectite (EXPANSIVE SOIL clay) content.

The list is fairly comprehensive, but it includes major and minor factors. It appears that there are probably two basic populations of sensitive clays: those with relatively low sensitivities, which contain clay minerals to a significant degree; and those with high sensitivities which tend to lack clay minerals. These high sensitivity materials are dominated by primary minerals with particle sizes in the clay or very fine silt ranges.

References

Mitchell, J.K. and Houston, W.N. (1969) Causes of clay sensitivity, *Journal of Soil Mechanics, American Society of Civil Engineers* 86(SM3), 19–52.
Quigley, R.M. (1979) Geology, mineralogy and geochemistry of soft soils and their relationship to geotechnical problems, *Proceedings of the 32nd Canadian Geotechnical Conference, Quebec City; State of the Art.*
Skempton, A. and Northey, R.D. (1952) The sensitivity of clays, *Géotechnique* 3, 30–53.

Further reading

Brand, E.W. and Brenner, R.P. (eds) (1981) *Soft Clay Engineering*, Amsterdam: Elsevier.

SEE ALSO: quickclay

IAN SMALLEY

SERPULID REEF

Serpulid reefs are built by marine polychaetes which secrete hard, calcareous tubes. The reefs built by *Ficopomatus enigmaticus* (Fauvel) develop reefs which are greater than 3 m thick and up to 20 m in diameter (Fornós *et al.* 1997). Individual worm tubes are typically 100 μm thick, 4–5 mm in diameter and up to 100 mm long, with tubes being comprised of calcium carbonate interspersed with a mucopolysaccharide matrix (Rouse and Pleijel 2001). Serpulidae have a near worldwide distribution; reefs have been found in the fossil record from the Cretaceous to the Recent and they are important in palaeoenvironmental reconstruction. The global distribution of the species has increased during historical times, due to international shipping, with several species colonizing non-native habitats and industrial structures such as docks.

The reefs typically develop in brackish and marine conditions, in intertidal and shallow subtidal zones, on hard rock substrata or shell fragments. Occasionally they extend onto, and in rare circumstances they colonize, soft substrates such as mud, sand or wood. Serpulid reefs tend to develop in three phases: (1) individuals coat the surface of hard substrata; (2) sinuous, random growth and extension of colonies; and (3) development of reef structures which are primarily controlled by water turbulence and dominant current direction. Morphological forms vary from fringing reefs along rocky shorelines, to subtidal microatolls and dense patch reefs (Fornós *et al.* 1997). Where serpulid reefs develop on soft substrata they often serve an important functional role, by providing the only hard substrate for a range of species.

References

Fornós, J.J., Forteza, V. and Martínez-Taberner, A. (1997) Modern polychaete reefs in Western Mediterranean lagoons: *Ficopomatus enigmaticus* (Fauvel) in the Albufera of Menorca, Balearic Islands, *Palaeogeography, Palaeoclimatology, Palaeoecology* 128, 175–186.

Rouse, G.W. and Plejil, F. (2001) *Polychaetes*, Oxford: Oxford University Press.

LARISSA NAYLOR

SHEAR AND SHEAR SURFACE

Shearing of material occurs when it is subjected to a stress or confining pressure, acting in a particular direction, that exceeds the strength of the material. The material fractures along a plane of least resistance, which may be curved, and the mass on one side of this shear surface is displaced so that movement takes place in opposite directions on either side.

In geomorphology, shear failure of materials is most commonly encountered in MASS MOVEMENTS. In a hillslope, material near the ground surface (soil or other sediments, *in situ* weathering products, bedrock, etc.) is subjected to shear stress as it is acted on by a downslope component of gravitational force. This is resisted by the shear strength of the material, which acts in the opposite direction. If the stress exceeds the strength, a shear surface will form (within one material or between different materials depending on relative strengths across all possible planes of weakness) and the mass above the surface will move downslope. (See HILLSLOPE, PROCESS; LANDSLIDE; SLOPE STABILITY.) Fault planes within the Earth's crust can also be regarded as shear surfaces: the rocks on one side of a fault move past the rocks on the other side in response to differential stresses.

Another condition that can cause shearing is when a material is confined at depth by stresses acting in perpendicular (x, y, z) directions, and compression by the stress in one direction exceeds the strength of the material. A shear surface forms as the material is forced to the side in opposite directions either side of a plane of least resistance.

The nature of a shear surface and any displacement along it will be determined by the properties of the material(s), the magnitudes and directions of the stresses, the topographic or geophysical context, and the scale at which the shear surface is considered. A smooth plane of movement at a large scale may comprise an irregular shear surface at a smaller scale. There may be considerable frictional resistance to movement, especially if the irregularities resemble interlocking teeth. At this extreme, the shear stress will only cause movement if it is sufficient either to force the two sides far enough apart to enable movement, or to cause shearing of the asperities thus further smoothing the overall surface.

At the other extreme, very small stresses within fine-grained materials may cause individual plate-like clay mineral particles to line up with the prevailing stress direction. Over time these 'microshears' can extend, join up and form a shear surface, which is smooth at the microscopic scale (see SLICKENSIDE).

Some materials will not form a discrete shear surface, instead developing shear zones where complex displacements occur throughout measurable thicknesses, especially in materials that display plastic behaviour under stress such as certain clay formations, or highly heterogeneous sediments.

Further reading

Selby, M.J. (1993) *Hillslope Materials and Processes, 2nd Edition*, Oxford: Oxford University Press.

SEE ALSO: riedel shear

ALAN P. DYKES

SHEET EROSION, SHEET FLOW, SHEET WASH

In a general sense, the term sheet flow is used to refer to any flow of water of more or less homogeneous depth and velocity over a surface without any clear channel development. The term is used to describe flow over alluvial floodplains, coastal plains, beach surfaces as well as flow over hillslopes without noticeable channel incision. Here, we will only refer to the use of the term sheet flow in this latter sense. Sheet flow over hillslopes therefore contrasts with flow in clearly defined channels (RILLs and gullies (see GULLY)).

Sheet erosion is often defined as the removal of a thin layer of surface soil by water erosion processes without the development of noticeable channels (Plate 121). The absence of channels (see CHANNEL, ALLUVIAL) does not imply that sheet erosion is completely uniform. The presence of roughness elements such as rock fragments, clods or vegetation tussocks will still lead to variations in flow depth and velocity, even under steady flow conditions. In order to stress the absence of channels the term interrill flow (i.e. flow on areas between or without rills) is sometimes preferred. Local variations in flow depth and velocity do not necessarily lead to channel development: as Smith and Bretherton (1972) pointed out, no channels will be formed by OVERLAND FLOW when the influx of sediment into a concentrated flowline by processes such as splash and creep equals or exceeds the local transporting capacity of the flow.

Sheet erosion mainly occurs under conditions where the soil surface is insufficiently protected by vegetation cover against drop impact and crusting (see CRUSTING OF SOIL): it is therefore frequently encountered on freshly ploughed arable land, on overgrazed rangeland and on natural hillslopes in arid and semi-arid environments.

If a soil surface is only affected by sheet erosion this will result in a gradual lowering of the whole soil surface. In many landscapes, sheet erosion often occurs simultaneously with rill and gully erosion, with gully erosion dominating in HILLSLOPE HOLLOWs (concavities), while sheet and rill erosion mainly affect convex and rectilinear slope segments. When hillslope erosion mainly occurs in channels, the much more rapid lowering of the channel beds will in principle result in a dissection of the soil surface. This need not always be the case: on arable land, rills formed due to erosion are frequently obliterated by soil tillage. The repeated occurrence of sheet and rill erosion will in this case also lead to a gradual removal of a 'sheet' of topsoil over the whole field surface. In natural conditions rill beds often stabilize after a certain time period due to bed ARMOURING: rills then become conveyers of sediment eroded on interrill surfaces rather than important sediment

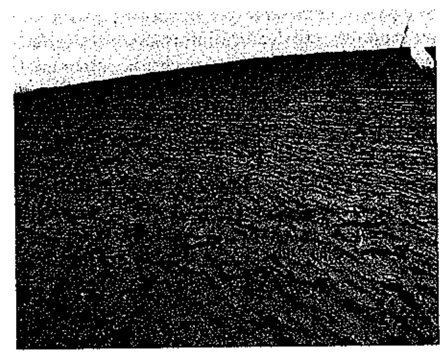


Plate 121 Intense sheet erosion on arable land in central Belgium, caused by a very intense rainstorm. Height of pole: *c.*1.5 m

sources. In the long term, this situation may also result in a gradual removal of the topsoil over the whole surface. The rest of this section concentrates on sheet erosion *sensu stricto*, i.e. water erosion that occurs without the formation of noticeable channels.

In the case of sheet erosion soil detachment mainly occurs by raindrop impact. The detached sediments may then be tranported by various processes: raindrop splash, raindrop-induced flow transport and flow transport. The hydraulic properties of the sheet (or interrill) flow are an important control on these processes and are therefore discussed first.

Sheet flow on hillslopes is generally characterized by flow depths between 0 and 20 mm and velocities $< 0.5\,\mathrm{m\,s^{-1}}$. In general, experimental studies in the field and laboratory show that classic approaches for the prediction of flow depths and velocities which are based on the estimation and prediction of the Darcy–Weisbach friction factor or Manning's n are only moderately successful (Abrahams *et al.* 1994). The key reason for this is that the interaction between roughness elements and flow hydraulics is fundamentally different in sheet flow. On natural surfaces where individual roughness elements are more or less randomly distributed, flow resistance tends to increase with increasing discharge until roughness elements are fully submerged. A further increase of discharge leads to a rapid decrease of flow resistance (Lawrence 1997). On tilled surfaces roughness elements are not randomly distributed and the correct prediction of flow characteristics sometimes requires separate calculations for each flowpath (Takken and Govers 2000). As the latter requires very detailed information on surface topography, simplified approaches are often used.

Raindrop detachment is basically controlled by the balance between rainfall erosivity and the soil's resistance to splash detachment. Various rainfall characteristics have been proposed as indicators of the rainfall's EROSIVITY. Rainfall kinetic energy is undoubtedly most frequently used to assess rainfall erosivity. However, there are indications that parameters that give relatively more weight to drop diameter, such as the product of rainfall momentum and drop diameter, are somewhat better predictors of splash detachment (Salles *et al.* 2000). Splash detachment is also strongly controlled by the presence and the thickness of a water layer on the surface: most studies report an exponential decline of splash detachment with water depth, with splash detachment becoming negligible when the depth of the water film exceeds *c.* one drop diameter (Torri *et al.* 1987). The resistance of a soil to splash detachment is determined by the cohesion of the soil and the weight of the soil grains. Minimum resistance values are observed for fine sandy soil materials as these materials consist of very fine grains yet they are almost totally cohesionless (Poesen 1985).

If a significant slope gradient is present, the redistribution of the detached material will result in a net downslope transport of soil. The rate of downslope transport increases approximately linearly with slope gradient. However, as splash distances are of the order of several decimetres, splash transport is generally negligible compared to (raindrop-induced) flow transport, except in the case of short, very steep slopes (e.g. unprotected terrace risers).

Although sheet flow has a limited capacity to detach (cohesive) sediment, it is in most cases the main transporting agent in sheet erosion. Two modes of sediment transport can be distinguished: (1) flow transport, whereby the flow itself is transporting the sediment and (2) raindrop-induced sediment transport whereby sediment is brought into the flow by raindrops impacting soil surface and the suspended sediment is consequently transported downslope. Raindrop-induced sediment transport flow is most important for relatively coarse (sand-sized) material and for low-energy flows with relatively low flow depths (Kinell 2001).

Sheet erosion is slope-dependent, primarily because the transporting capacity of the sheet flow increases with slope gradient. The increase of sheet erosion intensity with slope gradient is more or less linear, all other factors being constant. In practice other factors such as grain size distribution of the surface sediments, crusting and vegetation characteristics are strongly slope dependent, so that the relationship between slope and sheet erosion rates may take a completely different form. An interesting debate exists with respect to the effect of slope length on sheet erosion rates: recent experimental evidence suggests that the rapid increase of erosion rates per unit surface area which is often observed when plots of limited length are compared cannot be extrapolated to greater slope lengths. The increase of erosion rates with increasing slope length for short slopes is due to the fact that in the first

metres below the divide the sediment load in the sheet flow is supply-limited and therefore sediment transport increases rapidly with slope length. Further downslope, sediment transport becomes limited by the transporting capacity of the flow. This leads to a much slower increase of sediment transport with slope length and consequently also to a much slower increase or even a decrease of the erosion rate per unit area (Rejman and Usowicz 2002).

Considering the primary role of raindrop impact in detaching and transporting sediment in sheet flow, it is no surprise that sheet erosion is strongly dependent on rainfall and rainfall energy: this is, amongst others, reflected in the fact that rainfall erosivity in the Revised Universal Soil Loss Equation which is designed to describe both sheet and rill erosion equals the product of rainfall intensity and rainfall kinetic energy during a 30-minute period.

In general sheet erosion decreases strongly with increasing cover of the soil surface by vegetation or other non-erodible elements such as rock fragments. Most studies, such as the one by Hussein and Laflen (1982), report an exponential decline, whereby a soil cover of *c.* 30 per cent reduces erosion to less than 50 per cent of the value for a bare surface. This strong, non-linear response is due to the fact that the presence of cover has an impact on various aspects of the sheet erosion process (increased infiltration, protection of the surface cover against splash, reduced flow velocities, etc.). The adequate management of soil cover is therefore the most important management strategy in order to reduce sheet erosion.

References

Abrahams, A.D., Parsons, A.J. and Wainwright, J. (1994) Resistance to overland flow on semi-arid grassland and shrubland hillslopes, Walnut Gulch, Southern Arizona, *Journal of Hydrology* 156, 343–363.

Hussein, M.H. and Laflen, J.M. (1982) Effects of crop canopy and residue on rill and interrill soil erosion, *Transactions of the ASAE* 25, 1,310–1,315.

Kinnell, P.I.A. (2001) Particle travel distances and bed and sediment compositions associated with rain-impacted flows, *Earth Surface Processes and Landforms* 26, 749–758.

Lawrence, D.S.L. (1997) Macroscale surface roughness and frictional resistance in overland flow, *Earth Surface Processes and Landforms* 22, 365–382.

Poesen, J. (1985) An improved splash transport model, *Zeitschrift für Geomorphologie* 29, 193–211.

Rejman, J. and Usowicz, B. (2002) Evaluation of soil-loss contribution areas on loess soils in southeast Poland, *Earth Surface Processes and Landforms* 27, 1,415–1,424.

Salles, C., Poesen, J. and Govers, G. (2000) Statistical and physical analysis of soil detachment by raindrop impact: rain erosivity indices and threshold energy, *Water Resources Research* 36, 2,721–2,729.

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.

Takken, I. and Govers, G. (2000) Hydraulics of interrill overland flow on rough, bare soil surfaces, *Earth Surface Processes and Landforms* 25, 1,387–1,402.

Torri, D. Sfalanga M. and Del Sette, M. (1987) Splash detachment: runoff depth and soil cohesion, *Catena* 14, 149–155.

GERARD GOVERS

SHEETING

Some rocky massifs are divided by flat-lying or gently arcuate partings that at many sites are more inclined than the land surface, and plunge steeply (up to 70°). These fractures are known as sheeting or EXFOLIATION. Even though the term sheeting suggests thin layers, some are 10 m or more thick. Sheeting has been observed to depths of 100 m or more and is well developed in granitoids but also in other quite different rocks (dacite, rhyolite, sandstone, conglomerate and limestone). There are two interpretations of sheeting that fall into two categories: exogenetic and endogenetic. The exogenetic explanations are INSOLATION WEATHERING, CHEMICAL WEATHERING and offloading or pressure release, all of them imply dilation by rock volume increase.

The endogenetic explanations are all concerned with tectonic stresses. Some authors consider the sheet structure formed during the emplacement and later cooling of the granite mass and propose the same origin for the shape of the associated dome. But sheeting is well developed in sedimentary and volcanic rocks that were never emplaced and even in granites the magnetic foliation contemporaneous with the emplacement of magmatic rock is clearly discordant with the sheeting.

It has also been suggested that sheeting is associated with big thrust-shearing planes which would affect the granite as well as the other rock types where sheeting is habitually observed. This explanation is the best one, because it serves for all kinds of rocks affected by sheeting.

Further reading

Gilbert, G.K. (1904) Domes and dome structure of the High Sierra, *Geological Society of America Bulletin* 5 (15), 29–36.

Vidal-Romani, J.R. and Twidale, C.R. (1998) *Formas y paisajes graníticos*, Servicio de Publicaciones de la Universidad de Coruña, Serie Monografías 55.

Vidal-Romani, J.R. and Twidale, C.R. (1999) Sheet fractures, other stress forms and some engineering implications, *Geomorphology* 31(1–4), 13–27.

SEE ALSO: pressure release

JUAN RAMON VIDAL-ROMANI

SHIELD

The continental nuclei where Archean and Proterozoic rocks, typically burittle, rigid, granitic, gneissic and associated rocks, outcrop. They form extensive, flat, relatively stable areas (cratons) which have been relatively undisturbed since Precambrian time, except for gentle warping. The major shields are Canadian, Fennoscandian, Angaran (north-east Siberia), African, Brazilian, West Australian and East Antarctic. The shields are bordered by stable platforms, which are continental areas floored by shield extensions and covered by a relatively thin layer of sedimentary rocks.

Because they are ancient areas, they have often experienced bevelling of their rocks by erosion and are characterized by extensive erosional plains. Some shield areas have been fashioned in part by glacial erosion, as in Canada and on the Baltic Shield, but ancient saprolites may also be extensively developed (Lidmar-Bergström 1995). Because of the presence of limited relief and resistant rocks, shields tend to be areas with low rates of mechanical and chemical denudation (Millot *et al.* 2002).

References

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

Millot, R., Gaillardet, J., Dupré, B. and Allegré, C.J. (2002) The global control of silicate weathering rates and the coupling with physical erosion: new insights from rivers of the Candian Shield, *Earth and Planetary Science Letters* 96, 83–98.

A.S. GOUDIE

SHINGLE COAST

The term 'shingle' has been used for at least 400 years in Britain and some Commonwealth countries, to describe sediments composed of mainly rounded pebbles, larger in diameter than sand (>2 mm) but smaller than boulders (<200 mm). Elsewhere terms such as gravel, stone, levées de galets, playas de cantos, schotterwälle and steinstrand are used. A generalized world distribution of shingle coasts is given in Figure 148. In many locations shingle is mixed with sand, silt, clay or organic debris, resulting in a 'mixed' sediment beach (e.g. Kirk 1980), but all shingle and boulder beaches can be regarded as different types of 'coarse clastic' beach (Carter and Orford 1993).

In general shingle coasts have received less scientific attention than sandy and muddy shorelines. In part, this reflects the fact that, at a world scale, they are much less common. However, in recent decades there has been an increasing awareness of the geomorphologic, ecological and engineering significance of shingle coasts in the contexts of sea-level change, flood defence and habitat conservation. Such coasts are now recognized as an internationally important, but disappearing resource (Packham *et al.* 2001).

Shingle coasts form in WAVE-dominated locations where suitably sized material is available. At a global scale they dominate high latitudes and those areas of temperate shores which were affected by Quaternary glaciation. They are locally important in some other temperate and low latitude areas where high relief landscapes of suitable geology occur near the coast, near the estuaries (see ESTUARY) of high-energy rivers, or where coral (see CORL REEF) is present. Elsewhere they are of limited importance. Isla and Bujalesky (1993) describe shingle coast locations in Argentina, McKay and Terich (1992) and Forbes *et al.* (1995) in North America and Shulmeister and Kirk (1993) in New Zealand.

At a regional scale, lithographic composition determines shingle availability and durability. Hard materials such as flint, chert, granite, quartzite and some metamorphic materials survive much longer at this clast size than sandstones, limestones or shells. Around Great Britain some 19,000 km of shoreline have an important shingle component, with almost 3,500 km of these coasts being pure shingle (Sneddon and Randall 1993/1994). Many of the

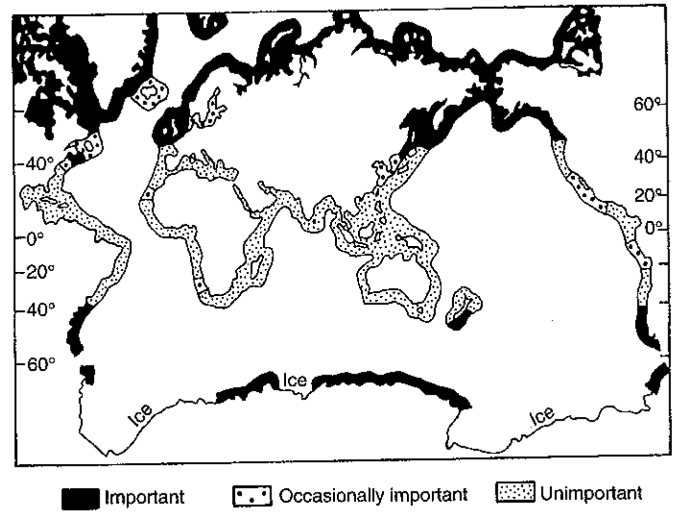


Figure 148 A generalized world distribution of shingle coasts (after Pye 2001)

shingle-barrier (see BARRIER AND BARRIER ISLAND) systems occurring on present-day coastlines were initiated during the Holocene (see HOLOCENE GEOMORPHOLOGY) marine transgression and are currently sustaining considerable morphological change as a result of increasing SEA LEVELS causing landward and longshore reworking of a finite sediment volume.

Shingle coasts can comprise several different landform types (Figure 149), which vary according to their history, mobility and oceanicity and therefore offer different habitats to vegetation, wildlife and humans (Pye 2001; Sneddon and Randall 1993/1994).

Fringing, or pocket BEACHes, are narrow strips of shingle coast in contact with the land along the top of the beach. These are usually subject to regular marine inundation. They frequently occur at the foot of sedimentary cliffs, such as chalk in southern Britain, but may also occur in front of coastal dunes or saltmarsh cliffs.

Embayment beach-ridge plains, or apposition beaches, are comprised of a series of relict storm beach-ridges and an active front ridge system which together partly or totally infill a previous embayment. Such systems may be hundreds of metres or even kilometres wide and can be transitional to CUSPATE FORELANDS or nesses.

Shingle SPITs are strips of shingle, which grow out from the coast where there is an abrupt change in the direction of the coastline. They commonly occur, therefore, along coasts which have an irregular plan. Spits often display recurved hooks along their length and at their distal ends, where the shingle is, or has been, subject to wave action from two or more directions. Indeed, in many cases, it is possible to trace the development of a spit's growth via recurved hooks, seen as lateral projections from the lee of the spit, which locate the position of the past distal points (Randall 1973; Plate 122). Paired spits are found at the entrance of several harbours on the south coast of England, including Pagham and Langstone. These may have originated as bars or TOMBOLOs, which have breached, but in other cases, independent growth of two spits may be due to bi-directional longshore drift.

On eroding coasts, shingle spits are transgressive and frequently overlay back-barrier marsh or lagoon deposits as at Shingle Street, Suffolk, and in some instances may be dissected to form barrier islands. Transgressive ridges, often composed mainly of shell-shingle, are well developed on the marsh coast of Essex. Similar features are also found in the Gulf of Mexico, where they are

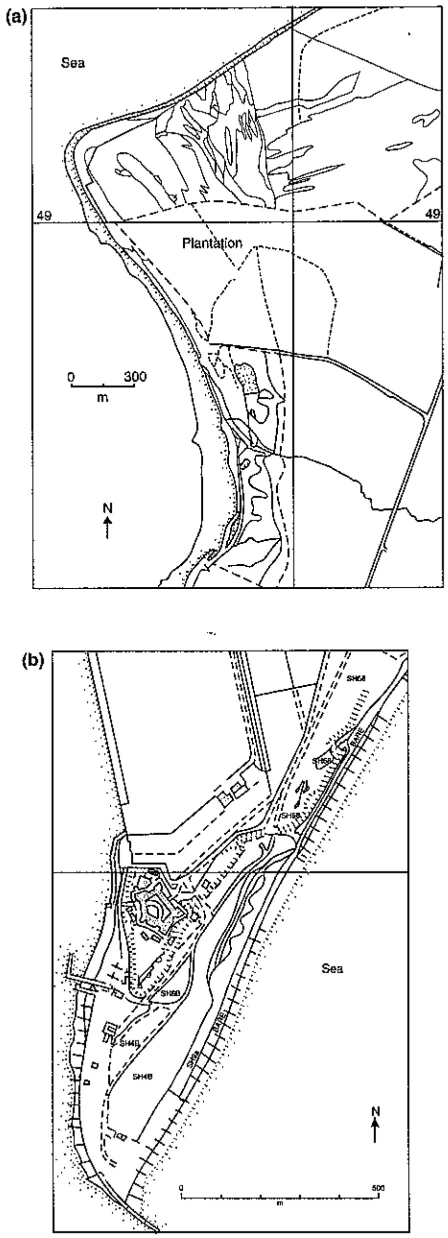


Figure 149 (a) A fringing beach at Llandulas in north Wales; (b) a shingle spit at Landguard Point, England; (c) a shingle bar at Culbin, Scotland; (d) an offshore barrier island, Scolt Head, England

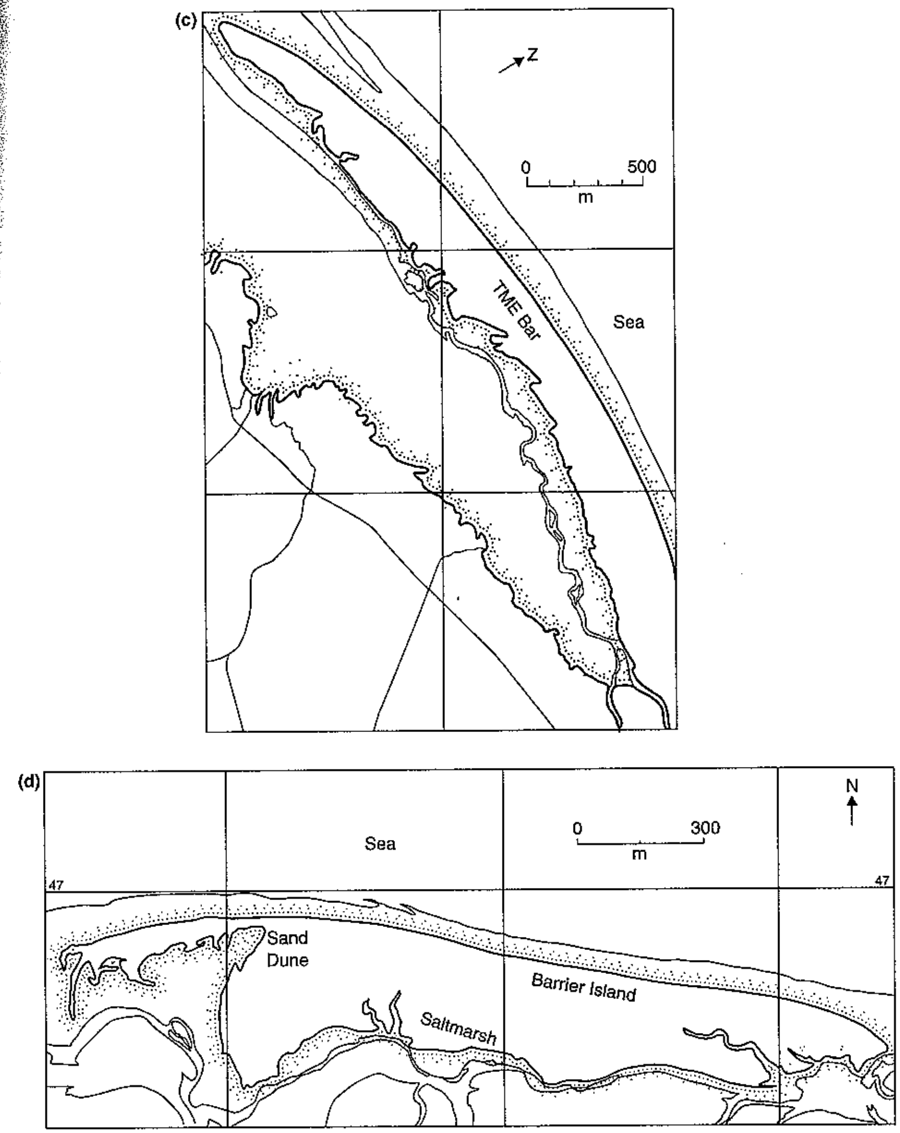


Figure 149 (Continued)

Plate 122 The shingle coast of Suffolk, UK. On the distal point of Orford Ness and on the mainland opposite there are recurved hooks enclosing lagoons. At the mouth of the estuary of the River Ore longshore drift of shingle sediments can be observed (photograph courtesy of Cambridge University Collection of Air Photographs)

known as CHENIER RIDGES, and in Auckland Bay, New Zealand. Tombolo barriers, or bars (see BAR, COASTAL), are geomorphologically similar to spits, representing the extreme case where a spit has grown across an estuary or coastal indentation. This results in the formation of a lagoon behind the bar, which clearly affects the hydrology and ecology of the leeward slope. Chesil Beach in Dorset and Slapton Ley in Devon are prime British examples.

Rivers, which provide a source of shingle-sized sediment (see GRAVEL-BED RIVERS), may have prograded strandplains or deltas of shingle at their mouth. In Scotland, the Kingston Shingles are found at the mouth of the Spey (Sneddon and Randall 1993/1994) and South Island, New Zealand has particularly good examples, such as at the mouth of the Waitaki River.

At points of littoral drift convergence the formation of a second set of apposition ridges deposited at a different angle, will lead to the formation of a ness or cuspate foreland, a triangular mass of shingle such as Dungeness, Kent, in England, Rhunahaorine Point, Argyll in Scotland or Cape Canaveral in Florida. The Island of Rügen in Baltic Germany is effectively a cuspate foreland cut off from the mainland. Such features often support a terrestrial geomorphic system inland of the coastal ridges.

The final type of shingle formation is the offshore barrier island, formed where a large mass of shingle has been deposited offshore and which may act as the 'skeleton' for a coastal sand-dune (see DUNE, COASTAL) system. Culbin Bar, Morayshire and Scolt Head Island, Norfolk, are prime British examples.

Most shingle coasts have a steep upper beach slope and a relatively steep overall nearshore profile. Partial wave energy reflection results in the formation of edge-waves and rhythmic longshore features such as BEACH CUSPS. Shingle features are frequently of considerable ecological importance in terms of habitat diversity and play a vital geomorphologic role in determining the stability of adjoining 'soft' sediments of mudflats and saltmarshes. Unless the shingle coastal features are mobile, a partial vegetation cover is the norm. The middle and lower beach are usually kept bare by wave action, but upper beaches are vegetated. The rate and extent of plant colonization is dependent upon the degree of disturbance and shingle mobility, the presence or otherwise of a fine sediment matrix within the spaces between larger sediments and the hydrological regime of the shingle.

All shingle coasts contain a mixture of different sized sediments. Some are well sorted and consist entirely of pebbles, while others are poorly sorted and may also contain sand and/or boulders. Because there is frequently considerable temporal and spatial variation in shingle and mixed shingle/sand beaches (Kirk 1980; Schulmeister and Kirk 1993), accurate determination of average textural qualities is difficult.

Most coarse sediment coasts become coarser up-beach, because backwash and gravity can move larger clasts. Hence many locations have shingle only on the upper beach. Williams and Caldwell (1988) also comment on clast shape with discoid pebbles sorted preferentially on the upper parts of the beach with spheres and rods occurring nearer the sea. Sediment grading alongshore also occurs due to selective transport of finer sediments in the downdrift direction as at Chesil Beach. However, other sites show much more complex patterns as a result of bi-directional currents of varying magnitudes.

Shingle coast micro-relief dynamics depend upon spring to neap tidal patterns and wind and wave conditions. The upper 50–80 cm of sediment is frequently remobilized forming berms and cusps that change from one tidal cycle to another. More major changes occur seasonally as a result of spring to neap tidal fluctuations and especially at those times when storm-wave energy is higher.

The internal sedimentary architecture of shingle landforms reflects the process regime and net evolutionary trends of the structure (Randall 1973). Ridge external structures vary dependent upon whether they are vertically accreting but laterally stable, laterally migrating or developing on a seaward prograding plain. The depressions between ridge crests may be partly filled by washover and storm-tossed deposits, so that there is often a marked difference in average particle size and shape between ridge fulls (crests) and lows (Randall and Fuller 2001). Sediment grading may also occur as a result of longshore drift with selective transport of finer sediments downdrift. However, on many coasts sediment grading has been found to be complex in relation to seasonal variation in the longshore current regime (Pye 2001).

Sea-level rise has the tendency to move shingle landforms inland (Carter and Orford 1993; Forbes *et al.* 1995), but if sea-level rise is particularly rapid, shingle structures may be drowned *in situ* by overstepping. Normally, however, under moderate storm-wave activity, shingle is pushed to the top of the front-beach ridge, while in major storms the ridge is overtopped or breached, creating shingle aprons in the backbarrier area. As this pattern is repeated, so the ridge migrates landwards by rollover. Many of the major shingle formations present today formed in this way during the Holocene marine transgression, initiating at a time of lower sea-stand and reaching their present location by around 4,000 BP. Most current shingle features are relict or dependent upon erosional sediments rather than glacial debris. Hence, there is currently a shortage of sediment at the updrift end of many transport cells and increasing risk of OVERWASHING and breaching (Orford *et al.* 2001).

Traditionally in developed countries, shingle coasts have been heavily managed to retard erosion, drift and sediment cycling and, more recently, for habitat conservation. Management methods may include beach reprofiling or protection (with gabions or tetrapods) or the construction of GROYNES and offshore breakwaters. Groynes have been used since the nineteenth century but frequently they have a negative effect downdrift by reducing longshore sediment availability. More recently beach nourishment has been seen as more cost effective and environmentally acceptable (Bradbury and Kidd 1998), but this, too, may change the natural geomorphologic character of the coast and may not be cost effective in the long term. Elsewhere gravel extraction, building developments, military activity and MANAGED RETREAT have markedly changed the landscape and landforms of shingle coasts.

At a world scale, large shingle structures are uncommon and under increasing pressure from development, mining and 'coastal squeeze' as a result of rising sea levels and coastal erosion. Most shingle structures were formed earlier in the Holocene Period and currently shingle supply is limited at the updrift end of coastal SEDIMENT CELLS. This results in the increased likelihood of breaching during STORM SURGES. Naturally, shingle coasts are dynamic and tend to migrate landward but people prefer a static coast. For economic reasons some areas of shingle coasts have to be protected, but in less developed areas space should be left for natural dynamic coastal change. Wherever possible shingle structures should be left entirely alone, since in the majority of circumstances, geomorphologic change promotes environmental diversity (Randall and Doody 1995).

References

Bradbury, A.P. and Kidd, R. (1998) *Hurst Spit Stabilisation Scheme – Design and Construction of Beach Recharge*, Proceedings of the 33rd MAFF Conference of River and Coastal Engineers, Keele University, July 1998, 1.1.1–1.1.13.

Carter, R.W.G. and Orford, J.D. (1993) The morphodynamics of coarse clastic beaches and barriers: a short and long term perspective, *Journal of Coastal Research* 15, 158–179.

Forbes, D.L., Orford, J.D., Carter, J.W.G., Shaw, J. and Jennings, S.C. (1995) Morphodynamic evolution self-organisation and instability of coarse-clastic barriers on paraglacial coasts, *Marine Geology* 126, 63–85.

Isla, F.I. and Bujalesky, G.G. (1993) Saltation on gravel beaches, Tierra del Fuego, Argentine, *Marine Geology* 115, 263–270.

Kirk, R.M. (1980) Mixed sand and gravel beaches: morphology, processes and sediments, *Progress in Physical Geography* 4, 189–210.

McKay, P. and Terich, R.A. (1992) Gravel barrier morphology, Olympic National Park, Washington State USA, *Journal of Coastal Research* 8, 813–829.

Orford, J.D., Jennings, S.C. and Forbes, D.L. (2001) Origin, development, reworking and breakdown of gravel-dominated coastal barriers in Atlantic Canada: future scenarios for the British Coast, in J.R. Packham *et al.* (eds) *Ecology and Geomorphology of Coastal Shingle*, 23–55, Otley: Westbury Publishing.

Packham, J. R., Randall, R.E., Barnes, R.S.K. and Neal, A. (eds) (2001) *Ecology and Geomorphology of Coastal Shingle*, Otley: Westbury Publishing.

Pye, K. (2001) The Nature and Geomorphology of Coastal Shingle, in J.R. Packham *et al.* (eds) *Ecology and Geomorphology of Coastal Shingle*, Otley: Westbury Publishing.
Randall, R.E. (1973) Shingle Street, Suffolk: an analysis of a geomorphic cycle, *Bulletin of the Geological Society of Norfolk* 24, 15–35.
Randall, R.E. and Doody, J.P. (1995) Habitat inventories and the European Habitats Directive: the example of shingle beaches, in M.J. Healy and J.P. Doody (eds) *Directions in European Coastal Management*, 19–36 Cardigan: Samara Publishing.
Randall, R.E. and Fuller, R.M. (2001) The Orford Shingles, Suffolk, UK: evolving solutions in coastline management, in J.R. Packham *et al.* (eds) *Ecology and Geomorphology of Coastal Shingle*, 242–260, Otley: Westbury Publishing.
Shulmeister, J. and Kirk, R.M. (1993) Evolution of a mixed sand and gravel barrier system in Canterbury, New Zealand, during the Holocene sea-level rise and still stand, *Sedimentary Geology* 87, 215–235.
Sneddon, P. and Randall, R.E. (1993/1994) *Shingle Survey of Great Britain: Final Report Appendix 1, Wales; Appendix 2, Scotland; Appendix 3, England*, Peterborough: Joint Nature Conservation Committee.
Williams, A.T. and Caldwell, N.E. (1988) Particle size and shape in pebble beach sedimentation, *Marine Geology* 82, 199–215.

SEE ALSO: coastal geomorphology; raised beach

ROLAND E. RANDALL

SHORE PLATFORM

Shore platforms are rock surfaces created by the erosion and retreat of coastal cliffs (Figure 150) (see CLIFF, COASTAL). Although geological and other factors are responsible for enormous variations in their morphology within fairly small areas, the distinction has often been made between subhorizontal, supra-, inter-, or subtidal platforms that terminate abruptly seawards in a low tide cliff, and gently sloping, largely intertidal, platforms, with gradients between about 1° and 5°, that continue below the low tidal level without a major break in slope (Plate 123). Subhorizontal platforms have generally been associated with Australasia, although they are common in much of the tropical and subtropical world, whereas sloping platforms have been described most frequently from the stormy waters of the northern Atlantic.

The inherent complexity of shore platforms and other rocky coastal systems has made it difficult to determine how they are formed or how they develop through time. The physical resistance of rocks depends upon their chemical composition, angle of dip, strike, bed thickness, joint (see JOINTING) pattern and density, degree of WEATHERING and a myriad of other factors. A wide range of mechanisms also operate on shore platforms, including WAVES, tides, bio-erosional and bioconstructional organisms, frost, chemical and salt weathering, WETTING AND DRYING WEATHERING and MASS MOVEMENTS. The relative and absolute importance of these processes have varied through time, with changes in relative SEA LEVEL and climate, and rock coasts often retain vestiges of environmental conditions that were quite different from today.

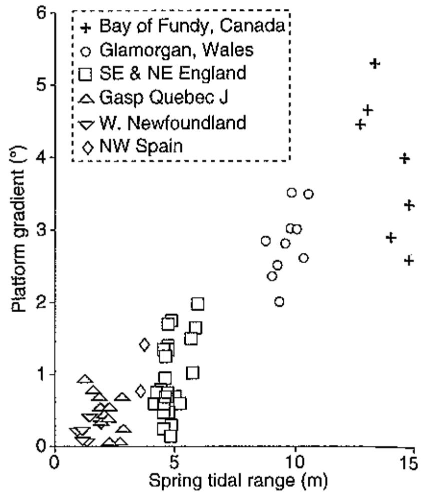


Figure 150 Shore platform in Liassic limestones and shales at Monknash, south Wales

Plate 123 Gently sloping and quasi-horizontal shore platforms

Much of the debate over the past one hundred years has been concerned with the relative roles of marine and subaerial processes in platform development, and, more recently, on the relationship between platform morphology and tidal range. A number of mechanisms have been proposed for platform formation:

- Platforms are cut in weathered or unweathered rock by waves. This produces sloping platforms (wave-cut platforms) in macrotidal regions and horizontal platforms, at the 'level of greatest wear' in areas with a small tidal range.
- Old Hat platforms develop in very sheltered areas, at the level of permanent seawater saturation. Above this level, weak waves wash away the fine, weathered debris, exposing the top of the resistant, unaltered rock below.
- Platforms can be formed by differential wave erosion of cliffs consisting of weak, weathered rocks above the saturation level, and more resistant, unweathered rock below. Evidence is lacking, however, for the existence of a permanent intertidal level of saturation in coastal rocks.
- Horizontal platforms, often with ridges or ramparts at their seaward ends, develop by WATER-LAYER WEATHERING and other weathering processes levelling and lowering uneven, rough, sloping or subhorizontal platforms that were originally cut by waves.
- Horizontal and sloping platforms are the result of alternate wetting and drying, which is responsible for cliff erosion and platform downwearing.
- Some workers have proposed that frost and possibly sea ice produces shore platforms in cool climatic regions.

It is increasingly evident that shore platforms are the product of mechanical wave erosion, weathering and bioerosional activity, although their relative importance depends upon climate, geology, wave and tidal conditions, and stage of development. Air compression in rock crevices and other mechanical wave erosional processes are most effective on steep uneven platform surfaces. As platforms become wider, smoother and more gently sloping, mechanical wave erosion must become less effective because of wave attenuation and the lack of rock scarps or upstanding beds of steeply dipping rock, and weathering processes must therefore become, at least relatively, more important. MICRO-EROSION METER (MEM) data from a variety of environments suggest that shore platforms are being lowered by weathering at rates that are frequently between about 0.5 to 1 mm yr^{-1}. It is difficult to accept that these high rates can be sustained indefinitely, however, and there is some MEM data to support the contention that they must decrease through time as platforms are reduced in elevation and therefore experience progressively longer periods of inundation, shorter periods of exposure and less frequent cycles of wetting and drying.

There is a moderately strong global relationship between mean regional platform gradient and tidal range (Figure 151). For wave-cut shore platforms this can be attributed to the way that tides control the expenditure of wave energy within the intertidal zone. The strong correlation between

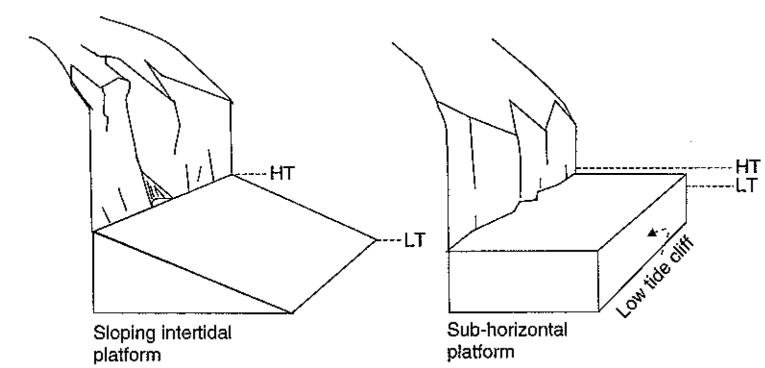


Figure 151 Relationship between mean regional shore platform gradient and spring tidal range

wetting and drying frequency distributions and tidal range also provides a possible explanation for the gradient–tidal range relationship in areas where weathering rather than wave action is dominant. Nevertheless, there is a basic problem with all theories that attribute platform formation entirely to subaerial or intertidal weathering, while relegating the role of waves to removal of fine-grained debris. This is concerned with the apparent lack of a mechanism to place a limit on maximum platform width, if wave strength and attenuation are not important factors.

Rock coast workers need to determine whether, or to what degree, shore platforms and related elements of rock coasts have been inherited from interglacial stages (see ICE AGES) when sea level was similar to today's. In hard, resistant rocks, the platforms often seem to be too wide to have developed in the few thousand years since the sea reached its present level, and they are frequently backed by ancient composite cliffs with glacial, periglacial or other terrestrial deposits, RAISED BEACHes and erosional ledges at their base. Although these coasts often lack datable materials, a variety of techniques has been used to show that at least in some areas, shore platforms, caves and other features have been inherited from one or more interglacial stage. It is particularly difficult to assess the possible role of INHERITANCE in areas of fairly weak rock, because platform width is less anomalous with regard to present rates of erosion than in more resistant rock areas, and because faster rates of erosion could account for the general lack of till covers, ancient beach deposits and structural remnants. Lacking evidence to the contrary, most workers have concluded that shore platforms in fairly weak rocks are entirely postglacial features. There is abundant evidence, however, of coastal inheritance from the last and older interglacial stages on the fairly weak rock coast of Galicia in northwestern Spain. Thick fluvio-nival and geliflucted slope deposits covered this coast during the latter part of the last glacial stage, and the ancient coast was then exhumed and inherited as sea level rose to its present position during the Holocene. The Galician evidence has important implications for the possible role of inheritance in the development of shore platforms in other areas (Trenhaile *et al.* 1999).

Modelling provides one of the few ways to study the long-term evolution of slowly changing rock coasts. The earliest MODELS were qualitative and structured within evolutionary cycles of erosion. More recent models are mathematical, but although field evidence suggests that most mechanical wave erosion occurs through water hammer, air compression in joints and ABRASION in shallow water, processes that are closely associated with the water surface, models have generally been concerned with submarine erosion in tideless seas. Nevertheless, a model has been developed that considers rates of wave attenuation and the long-term distribution of wave energy within the intertidal zone. This model has been used to study the long-term evolution of shore platforms with Quaternary changes in sea level on stable and tectonically mobile coasts. The model indicated that whether an ancient shore platform is subsequently inherited, modified or completely replaced by a contemporary platform depends upon the complex interaction of a multitude of factors that determine the erosive efficacy of the waves. It suggested that intertidal and subtidal surfaces trend towards a state of static equilibrium under oscillating sea level conditions, as attenuated waves become increasingly unable to continue eroding the rock, although they can be in a temporary, though possibly long-lasting, state of DYNAMIC EQUILIBRIUM. Most modelled surfaces were, at least in part, inherited from one, or in many cases more, interglacial stages when sea level was similar to today's (Trenhaile 2001). In future, platform models must also consider the effects of downwearing by weathering as well as backwearing by waves within the intertidal zone. In turn, reliable modelling is dependent on the acquisition of quantitative field data. Although the number of sites and the length of the records are quite limited, the micro-erosion meter has provided useful information on platform downwearing by weathering and corrasion. We are still unable, however, to measure the effect of joint block and large rock fragment detachment by wave quarrying and other mechanisms.

References

Trenhaile, A.S. (2001) Modelling the Quaternary evolution of shore platforms and erosional continental shelves, *Earth Surface Processes and Landforms* 26, 1,103–1,128.

Trenhaile, A.S., Pérez Alberti, A., Martínez Cortizas, A., Costa Casais, M. and Blanco Chao, R. (1999) Rock coast inheritance: an example from Galicia, northwestern Spain, *Earth Surface Processes and Landforms* 24, 605–621.

Further reading

Stephenson, W.J. (2000) Shore platforms: remain a neglected coastal feature', *Progress in Physical Geography* 24, 311–327.

Sunamura, T. (1992) *The Geomorphology of Rocky Coasts*, Chichester: Wiley.

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

SICHELWANNE

Crescent-shaped grooves formed by action of a glacier. The term is derived from Germany, meaning 'sickle tubs', but is part of a collective set of features known as plastically sculptured forms, or p-forms. Sichelwannen (plural) occur in crystalline rocks set in glacial environments, and range in size from 1–10 metres in length, 5–6 metres wide, and from millimetres up to several metres in depth. They are a transverse type of p-form, commonly displaying striations, and with the horns of the crescent shape pointing down-glacier. The origin of sichelwannen is unclear, and several methods of formation have been suggested in the literature. The most plausible manner of formation is erosion by high pressure subglacial meltwater. Similar forms to sichelwannen have been produced by water in less resistant rock, where up-slope topographic obstructions force subglacial channels to bifurcate, producing the characteristic horseshoe pattern on the stoss-side of the obstacle. Erosion by glacier ice has also been proposed, supported by the presence of striations on sichelwannen. However, advocates of a fluvial origin believe the distinctively patchy nature of the striations prohibits formation by ice.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*. London: Arnold.

Shaw, J. (1994) Hairpin erosional marks, horseshoe vortices and subglacial erosion, *Sedimentary Geology* 91, 269–283.

STEVE WARD

SILCRETE

A terrestrial geochemical sediment arising from low temperature near-surface physico-chemical processes operating within the zone of WEATHERING in which silica has accumulated in, and/or replaced, a pre-existing soil, sediment, rock or weathered material. Silcretes are defined as containing >85 per cent silica by weight, with some pure examples consisting of >95 per cent silica (Summerfield 1983). They can be distinguished from other silica-cemented sedimentary rocks such as orthoquartzites as they often exhibit a porphyroclastic, as opposed to even-grained, texture when viewed in microscopic thin-section (Watson and Nash 1997). Silcretes commonly consist of hard, silica-cemented quartz sand or brittle quartzitic material with a conchoidal fracture. The cement can consist of a range of silica minerals, of which opal, chalcedony, cryptocrystalline silica and quartz are the most widely documented. The presence of other minerals may affect the silcrete colour, with grey, brown and green varieties reported.

Although not as common as many other DURICRUSTs, silcretes are widely distributed in low latitude and other environments, particularly those areas which did not undergo Late Tertiary and Quaternary glaciation. They are found on every continent except Antarctica but are most widespread in inland and southern Australia, and in the Kalahari and Cape coastal regions of southern Africa. Other locations with significant areas of silcrete include Britain (where they are termed 'sarsens') Europe, the Sahara, Tanzania, USA, Uruguay and Brazil. Given their hardness and chemical stability, they are extremely resistant to erosion and often act as CAPROCKs and influence INVERTED RELIEF development. Silcretes most commonly occur as distinct horizons, but may also form a coating on rock outcrops or lenses within other duricrusts such as CALCRETE. Well-developed silcrete horizons are between 0.5 and 3 m thick, although thicknesses of >10 m have been recorded. A variety of terms have been used to describe profiles including massive, columnar, nodular, glaebular and mammilated, reflecting the numerous surface morphologies and modes of origin of many silcretes.

Silcrete development can take place via a variety of pedogenic and non-pedogenic processes, but all require a silica source, a means of transferring this silica to the site of formation, and a mechanism to trigger precipitation. The most significant source is CHEMICAL WEATHERING of silicate-rich rocks, particularly those containing clay minerals. Silica released in this way may then be available for transport in solution. Highly alkaline conditions, such as those found in arid zone lakes, can also lead to extensive dissolution of silicate minerals. Quartz is only weakly soluble in neutral pH terrestrial surface waters (around 10 ppm at 25 °C), but solubility increases dramatically above pH 9.0. Other silica sources include replacement of quartz during carbonate precipitation, dissolution of volcanic and other dust, and biological inputs from silica-rich plants and micro-organisms. Silica from these sources may be transferred in solution to the point of precipitation via lateral or vertical movements of ground water, pore water and surface water, with a range of local and far-travelled silica potentially contributing to silcrete formation. Silica precipitation may also be initiated by a variety of factors, of which the most important are evaporation, cooling, organic processes, absorption by solids, reactions with cations and changes in pH (particularly a shift to below pH 9.0 in alkaline environments).

Silcrete formation by pedogenic processes involves the accumulation of silica from downward percolating soil water during a succession of cycles of leaching and precipitation. Such silcretes commonly consist of a nodular base overlain by a columnar section containing illuviation structures, capped by a brecciated component. Silica mineralogy often varies throughout the profile, with more ordered forms of silica cement in the uppermost sections and less ordered forms towards the base (Milnes and Thiry 1992). Pedogenic silcretes may develop over large areas, are often semi-continuous and relate directly to palaeosurfaces. Non-pedogenic models embrace formation in a variety of settings, including zones of water table fluctuation or groundwater outflow, locations marginal to drainage lines, as well as lakes and seasonal pools. Silica precipitation in these environments is usually controlled by evaporation, pH shifts or organic processes, with the resulting silcretes forming localized sheets or lenses. Non-pedogenic silcretes are usually simple in terms of their macro- and micromorphology, although a range of silica cements may be present. Significantly, they are almost always devoid of the organized profile associated with pedogenic silcretes. Non-pedogenic silcretes may also form at considerably greater depths than pedogenic types and therefore do not normally represent a palaeosurface.

Perhaps the greatest controversy surrounding silcrete is the degree to which environmental controls determine formation, a critical issue if silcretes are to be used as palaeoenvironmental indicators (Nash *et al.* 1994). Silcrete has been suggested to form under climates ranging from semi-arid to monsoonal on the basis of geochemical, mineralogical and micromorphological evidence, and by comparison of the geographic and stratigraphic distribution of silcrete with contemporary and past climate. The situation is made more complex by the fact that most silcretes are relict. Contemporary silcrete formation has only been documented from one location, a hypersaline lake in the Kalahari (Shaw *et al.* 1990). Unfortunately, silicification at this site is driven by organic silica fixation so the resultant silcretes are not an ideal modern analogue. The only safe conclusion that can be made is that the presence of silcrete should not be considered diagnostic of specific environmental conditions. It is essential to establish the mode of origin of any silcrete and view formation within the context of other climate proxies before using it as a palaeoenvironmental indicator. Pedogenic silcretes, such as those in southern South Africa, may have taken hundreds of thousands of years to form and, as a result, have integrated climatic effects over considerable time periods. In contrast, some non-pedogenic groundwater silcretes, such as those in the Paris Basin, developed in a few tens of thousands of years under steady groundwater outflow.

References

Milnes, A.R. and Thiry, M. (1992) Silcretes, in I.P. Martini and W. Chesworth (eds) *Weathering, Soils and Palaeosols*, 349–377, Amsterdam: Elsevier.

Nash, D.J., Thomas, D.S.G. and Shaw, P.A. (1994) Siliceous duricrusts as palaeoclimatic indicators: evidence from the Kalahari Desert of Botswana, *Palaeogeography, Palaeoclimatology, Palaeoecology* 112, 279–295.

Shaw, P.A., Cooke, H.J. and Perry, C.C. (1990) Microbialitic silcretes in highly alkaline environments: some observations from Sua Pan, Botswana, *South African Journal of Geology* 93, 803–808.

Summerfield, M.A. (1983) Silcrete, in A.S. Goudie and K. Pye (eds) *Chemical Sediments and Geomorphology*, 59–91, London: Academic Press.

Watson, A. and Nash, D.J. (1997) Desert crusts and varnishes, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 69–107, Chichester: Wiley.

DAVID J. NASH

SINGING SAND

Two types of sand that emit audible sounds with a coherent wave pattern, on being sheared by wind or other mechanical means, have been reported throughout the world over the past century. They are known as squeaking or singing (found on beaches) in booming (found in sand dunes) sands. The exact mechanism by which these coherent sounds are produced from these sand materials is still unknown.

Many acoustical sands are well-sorted materials with sizes in the range of 100–500 microns, roughly rounded particles with a high content of quartz particles. However, booming sand materials are also found in Kauai, Hawaii with a high calcareous content.

Various material science techniques have been used to show that there is a thin rind-like layer on the particles of these acoustical sand materials. Direct Scanning Electron Microscopic examination of particles that have been sliced by grinding show this layer in both pure silica-gel singing materials and the calcareous Kauai sands. In the former case, the rind layer is composed of amorphous silica and in the latter, an alumino-silicate clay-like material. The rind width is about 5 microns.

Fourier Transform Infra Red (FTIR) analysis has shown that acoustical sand exhibit a broad absorption band in the range of 2,800 cm^{-1}. This band is due to clusters of water in the rind-like layer.

Etching of acoustical sands with hydrofluoric acid (HF) removes the surface layer and causes the sand to become silent. In the case of the Kauai sands, sonication is sufficient to remove this layer and this technique also silences the booming property.

Sand materials that do not sing can be made to do so by grinding the grains in a mill, periodically removing the fines and renewing the water, leaving a well-sorted and highly polished material which 'sings' and exhibits the characteristic FTIR 3,400 cm^{-1} broad band.

Finally, excess heating of these materials also removes this singing or booming sound.

Interest in finding the underlying mechanism of the production of this coherent wave phenomenon in granular materials is due to its possible use in a saser device for sonic hammering.

Further reading

Goldsack, D.E., Leach, M.F. and Kilkenny, C. (1997) Natural and artificial 'Singing Sands', *Nature* 386, 29.

DOUGLAS GOLDSACK AND MARCEL LEACH

SINUOSITY

Few natural rivers are straight for more than a few channel widths. Rivers that are not completely straight are sinuous, even if they are not clearly MEANDERING in the sense of having more or less regular oscillations in direction. Schumm (1963) introduced a quantitative definition of sinuosity as the distance along a river between two points A and B, divided by the valley distance between A and B. It is therefore a dimensionless ratio with a minimum value of 1.0 and a realistic maximum of around 3 to 4, after which neck cut-offs occur. It can be calculated for a single bend (when A and B are successive crossovers), a series of bends, or a longer reach.

The sinuosity of a modern channel, or a well-preserved palaeochannel, is readily determined from a map or aerial photograph. Some ambiguities of operational definition must be kept in mind when comparing values quoted by different authors. Is valley distance defined as a straight line, or does it allow for large-scale valley bends? If the river is braided (see BRAIDED RIVER), is the sinuosity computed for the centreline, the biggest channel, or the sum of all the channels as suggested by Richards (1982)?

Sinuosity can alternatively be assessed using variograms and FRACTAL concepts (Nikora 1991; Lancaster and Bras 2002), which can reveal any scale dependence as one moves from single simple bends to compound loops and multiple loops. The sinuosity of a fragmentary palaeochannel can be estimated from the variance of channel direction at places where this can be reconstructed (Ferguson 1977).

Evidently sinuosity varies spatially according to the straight, meandering, or other channel pattern of different reaches. Sinuosity can also fluctuate over years or decades as bends of an actively meandering channel grow and are cut off, and it may change progressively in the event of climate change, flow regulation, or other disturbance of the system.

Sinuosity has significance for fluvial processes and channel regime because it can be written not as a ratio of distances but of slopes: the valley slope divided by the channel slope. Since valley slope is largely inherited, an increase in the meandering tendency of a river leads not only to greater sinuosity but also reduced channel slope. This has consequences for velocity, shear stress, and BEDLOAD transport. Sinuosity is therefore regarded by many geomorphologists and other fluvial scientists as one of several channel properties that can adjust if the bedload supply to a reach is not the same as the transport capacity.

References

Ferguson, R.I. (1977) Meander sinuosity and direction variance, *Geological Society of America Bulletin* 88, 212–214.
Lancaster, S.T. and Bras, R.L. (2002) A simple model of river meandering and its comparison to natural channels, *Hydrological Processes* 16, 1–26.
Nikora, V.I. (1991) Fractal structures of river plan forms, *Water Resources Research* 27, 1,327–1,333.
Richards, K. (1982) *Rivers: Form and Process in Alluvial Channels*, London: Methuen.
Schumm, S.A. (1963) Sinuosity of alluvial rivers on the Great Plains, *Geological Society of America Bulletin* 74, 1,089–1,100.

ROB FERGUSON

SKERRY

A term that describes the low rocky islets common to many mid and high latitude coastlines. Many skerries may be covered at high tide and are subject to wave processes with the result that they may carry the signature of marine quarrying and abrasion. However, in spite of a marine influence, many skerries have also been shaped by past glacial erosion and, within the constraints of the geological structure of the host rock, may have inherited a moulded bedform or even roche moutonée shape. For example, many skerries in Finland, Sweden and Norway, particularly in sites sheltered from severe wave activity, still retain striations etched onto glacially moulded bedforms. This trait can also be seen in the rocky islets of formerly glaciated lakes such as in Loch Lomond in Scotland and in Lake Nasijarvi in Finland. On the coast, fields or chains of skerries frequently occur offshore of areally scoured surfaces that have been partly submerged by Holocene sea-level rise. Good examples occur in the Outer Hebrides of Scotland and in arctic Canada, and along the STRANDFLAT coasts of Norway, Sweden and Finland in the Baltic and in western Iceland.

Further reading

Bird, E.C.F. and Schwartz, M.L. (eds) (1985) *The World's Coastline*, New York: Van Nostrand Reinhold.

JIM HANSOM

SLAKING

The disintegration of a loosely consolidated material following the introduction of water or exposure to the atmosphere (Plate 124). Clays and shales (mudrocks) are especially prone to this form of failure, especially in the presence of saline waters. Materials with high Exchangeable Sodium Percentages (ESP), including some colluvia, may be susceptible, and slaking is an important process on many badland surfaces, including DONGAS. Various tests are available for determining the durability of slaking-prone materials (Czerewko and Cripps 2001) (see WETTING AND DRYING WEATHERING).

Reference

Czerewko, M.A. and Cripps, J.C. (2001) Assessing the durability of mudrocks using the modified jar slake index test, *Quarterly Journal of Engineering Geology and Hydrology* 34, 153–163.

A.S. GOUDIE

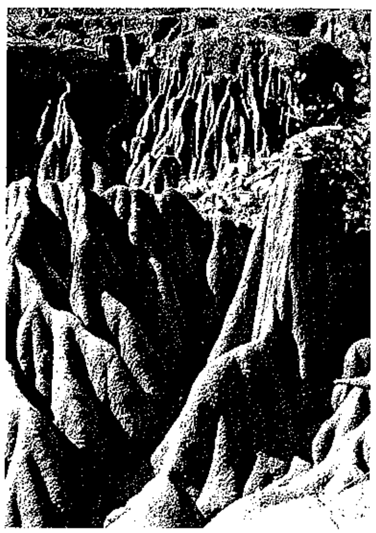


Plate 124 Materials with a high Exchangeable Sodium Percentage, such as this colluvial deposit from central Swaziland, are prone to severe erosion as a result of their propensity to slaking following rain events

SLICKENSIDE

A polished, striated rock surface on a fault or bedding plane caused by the frictional movement between one rock mass sliding over another. Slickensides are a common feature on fault planes, and though sometimes can be featureless, they commonly display prominent parallel ribbing or striation. These striations may be exhibited on mineral coatings, such as quartz and calcite, as well as on the rock itself, and can provide an indication of the direction of fault movement from their orientation (they form parallel to the direction of fault displacement). However, striations can often be erased by subsequent fault movement, and thus should only be considered as a record of the most recent fault movement. Additionally, from analysis of a suite of slickenside striations, an estimation of the magnitude of the *in situ* stress field can be established. The origin of slickenside striations is uncertain. Some may be grooves formed by outcropped resistant rock on the opposite fault block, or mineral lineations that grow with their long axis parallel to the direction of fault movement. Slickensides with striations often contain small steps oriented in one direction similarly to the striations.

The term slickenside also refers to natural crack surfaces along planes of weakness in soils, resulting from the movement of one mass of soil against another. This is commonly by the swelling and contraction of soils with high clay levels able to swell (e.g. montmorillonite). Slickenside also refers to the polished surface produced by the passage of a mudflow.

Further reading

Doblas, M. (1998) Slickenside kinematic indicators, *Tectonophysics* 295, 187–197.
Tjia, H.D. (1964) Slickensides and fault movements, *Geological Society of America Bulletin* 75, 638–686.

SEE ALSO: fault and fault scarp

STEVE WARD

SLOPE, EVOLUTION

Landscapes change over time in response to the internal redistribution of sediment, usually with some net export of material to rivers or the ocean. The way in which landscapes and slopes evolve depends on their initial form, the slope processes (see HILLSLOPE, PROCESS) operating and the boundary conditions which determine where and how much sediment is removed. This discussion will mainly focus on two-dimensional slope profiles, but some of the aspects which can only be addressed in 3-D will also be discussed below.

Slope evolution can be described in conceptual development sequences, and much of the history of early twentieth-century geomorphology was concerned with championing alternative conceptual models (Chorley *et al.* 1973), under the banners of W.M. Davis, Walther Penck and others (see CYCLE OF EROSION). More recent approaches have focused increasingly on the application of MODELS in geomorphology, and it is instructive to compare the various development sequences in these terms, to understand the conditions under which each is most appropriate.

Although G.K. Gilbert was not primarily a theoretician, his work in the Henry Mountains (1877) repeatedly interpreted slope profiles in the context of the interaction between form and process, introducing the term DYNAMIC EQUILIBRIUM to describe this balance. He correctly attributed the convexity of divides to SOIL CREEP and similar diffusive processes; and the concavity of the lower slopes to SOIL EROSION processes.

W.M. Davis (1909) spent much of his life canvassing the concept of the Geographical Cycle, which described what he perceived as the 'normal' sequence of erosional landforms, strongly based on his experience of humid temperate conditions. The cycle assumed the rapid uplift of a low relief landscape, and its erosion during a period of tectonically stable conditions. Generally the landscape is taken as soil-mantled, with three life stages. In *youth*, rivers incise deeply into the landscape, and hillsides gradually encroach upon the original surface, parts of which may survive as an *erosion surface*. Slopes become *mature* when the original surface has been consumed, and the highest points begin to undergo appreciable erosion. During maturity, slopes decline in gradient everywhere, forming a connected or *graded* system conveying material to the lowest point. Eventually maturity gives way to *old age* as the surface is reduced to a low relief surface, or PENEPLAIN, which may still retain a few remnant hills, or *monadnocks*. This sequence of ever-reducing relief could be interrupted by relative falls in sea level, which might *rejuvenate* the landscape, perhaps creating flights of partial erosion surfaces which preserved the morphology of previous interrupted cycles. Davis, and his disciples like Johnson and Wooldridge, used the methodology of the geographical cycle to reconstruct former sea levels and the history of the landscape from the inferred remains of high level erosion surfaces and sequences of river and coastal terraces. The two key assumptions of Davis's approach were first that uplift was rapid and followed by long periods of tectonic stability, and second that hillsides were essentially soil-mantled.

Walther Penck (1953 [1924]), working in the much more tectonically active area of the Andes, considered that slope forms responded primarily to the rate of uplift, which he assumed to be a continuous, if episodic, process. In perhaps his central conceptual model, he considered that convex slopes were produced where the rate of uplift was accelerating, and concave slopes where uplift rates were falling. The normal situation, of steady uniform uplift, was associated with slopes of uniform gradient, and these would retreat parallel to themselves while the uplift continued. This approach differs from Davis's in two important respects: first in the assumption of active tectonics, and second in coupling evolution of the slope to conditions at the slope base. Both these assumptions are most relevant where slopes have little or no regolith cover, and all detached material is immediately removed.

Lester King (1953), working in the semi-arid climate of South Africa, proposed a morphologically intermediate conceptual model, in which the steep and rocky upper slopes retreat parallel to themselves, and undergo replacement by a PEDIMENT, which is a lower gradient surface with some regolith cover. Pediments gradually coalesce and eventually consume any residual mountain masses. The combination of rocky and regolith-covered slope elements is significant in this model, while the tectonic assumptions return to the stability advocated by Davis.

In comparing these conceptual models, one of the important distinctions concerns the presence or absence of regolith on the slopes. Where there is a deep regolith, slope processes are able to act at their full capacity, and removal is said to be transport-limited or flux limited. Where the regolith is thin or absent and slopes are steep, material is removed as fast as it is detached by weathering or entrainment, and removal is said to be weathering-limited or supply limited (see WEATHERING-LIMITED AND TRANSPORT-LIMITED). In this case the actual transport rate is well below the *transporting capacity* of the sediment transport processes. Slope evolution is radically different according to which of these two regimes is active, and there is the possibility of an intermediate regime, in which both detachment rates and transport capacities are important. The nature of the regime can be seen from the *travel distance* of material during a transport event. Where travel distance is short compared to the slope length, for example under soil creep, rainsplash, bedload transport or soil erosion, removal is transport-limited; whereas where travel distance is long, for example under rock fall, debris flows, washload transport or movement of solutes, removal is essentially weathering-limited.

Several conceptual models have also been proposed to describe the conditions and forms associated with equilibrium or quasi-equilibrium forms for slope profiles. The essence of these approaches is that the profile form is considered to be independent of the initial surface form which is eroded. J.T. Hack (1960) has been associated with a particular form of DYNAMIC EQUILIBRIUM, in which the landform remains essentially fixed in form as it erodes. This form may strictly occur only where balanced by equal and opposite tectonic uplift, but Hack argued that, during the stage of Davisian maturity, much of the Appalachian landscape eroded with little change in form except close to base level, so that dynamic equilibrium provided a working approximation to observed conditions. M.J. Kirkby (1971) proposed the concept of *characteristic forms*, in which slope profiles retain their form, with more and more subdued relief as they decline in elevation, corresponding to the transition from Davisian maturity to old age. This is seen as a quasi-equilibrium in which the profile form is characteristic of the ensemble of processes acting on it.

By comparing these qualitative concepts with simple slope models, the conditions for all these simplified forms can be compared in a quantitative way. For a slope profile, evolution is controlled by four sets of constraints. First, mass must be conserved; second, evolution takes place from an initial profile form; third, evolution is subject to boundary conditions, which define, for example, conditions at the top (divide) and bottom (stream) of a hillslope; and fourth, evolution occurs through sediment transport by one or, usually, more slope processes. The transporting capacity for each process varies in some way with topography, usually with distance from the divide and gradient; and removal is also subject to transport or weathering-limited conditions, which can be defined by a fuller specification of the slope processes.

The Mass Balance equation for a slope profile states that:

Input − Output = net increase in Storage

or in the simplest case of a simple slope profile from divide to stream:

$$\frac{\partial z}{\partial t} + \frac{\partial S}{\partial x} = 0 \qquad (1)$$

where S is sediment transport per unit width, z is elevation above an arbitrary datum, x is horizontal distance measured from the divide, and t is elapsed time.

The cases of weathering and transport-limited removal can be distinguished using a sedimentation equation:

$$\frac{dS}{dx} = D - \frac{S}{h} \qquad (2)$$

where D is the rate of sediment detachment, and h is the travel distance (the mean distance travelled by sediment in an event). The second term of the right-hand side is the rate of deposition. Where sediment detachment balances sedimentation, the flow is said to be at its travel capacity, $C = D \cdot h$.

Where the travel distance is small (in relation to the slope length), then removal is essentially transport-limited and $S = C$, so that equation (1) can be used to define slope evolution, with C replacing S. Where the travel distance is long ($h \gg x_0$), removal is weathering-limited and slope evolution is approximated by equation (2) with the second term on the right-hand side negligible. Between these extremes, it is necessary to retain both equations (1) and (2) to determine how slopes evolve.

The simplest boundary conditions are of a summit fixed in horizontal position (at $x = 0$), and a stream fixed at some point $x = x_0$.

For several processes, capacity sediment transport rates can be written in the form:

$$C \propto x^m \Lambda^n \qquad (3)$$

For example simple but empirically acceptable formulations are shown in Table 43, although EQUIFINALITY allows some range of possible exponent values. Mass movements may also be included in a similar scheme, but with two threshold gradients: a lower threshold below which no movement occurs, and an upper threshold above which material will never stop. The lower threshold Λ_T corresponds to the maximum stable slope gradient under saturated conditions, and the upper threshold Λ_0 to the angle of repose for debris. A simple formulation then takes the form:

$$D \propto \Lambda(\Lambda - \Lambda_T);\ h \propto 1/(\Lambda_0 - \Lambda);$$
$$C \propto \Lambda(\Lambda - \Lambda_T)/(\Lambda_0 - \Lambda) \qquad (4)$$

with the capacity, C, defined only in the range $\Lambda_T < \Lambda < \Lambda_0$, and travel distances generally of the order of the total slope length. By treating mass movement as a continuous process, the feedbacks produced by the size of an individual slide event are ignored, so that this formulation works best for small and shallow slides, and for rockfalls from cliffs.

Without entering into a full analysis of these equations, some results can be quoted here without proof. First, the conditions for downcutting at a steady uniform rate, T corresponding to a strict application of Hack's dynamic equilibrium, is that:

$$S = Tx \qquad (5)$$

This is necessarily true because, in the steady state, the slope processes must carry away all the material eroded upslope.

For any transport-limited removal process (i.e. S = C), equation (3) gives:

$$C = Tx \sim x^m \Lambda^n, \quad \text{or} \quad \Lambda \sim Tx^{(1-m)/n} \qquad (6)$$

Thus, for steady-state downcutting, hillslopes are convex (gradient increasing downslope) when the distance exponent $m < 1$, and concave if $m > 1$. This means that slopes become convex for soil creep, rainsplash, solifluction and tillage erosion, and become concave for rillwash. More realistically, a combination of processes is acting, for

Table 43 Exponents for some sediment transport processes in equation (3)

Process	Travel distance	m	n
Soil creep	Small	0	1
Rainsplash	(<1 m)	0	1
Solifluction	(<1 m)	0	1
Tillage erosion	(<1 m)	0	1
Rillwash	~10 m	2	1–2
Solution	≫1 km	1	0

example rainsplash and rillwash. Adding these terms the combined transporting capacity may be written as:

$$C = k\Lambda\left[1+\left(\frac{x}{u}\right)^2\right] \tag{7}$$

where u is the distance beyond which the rillwash term (the second term on the right-hand side) becomes greater than the rainsplash term (first term on RHS). Thus for $x<u$, rainsplash is the dominant process, and for $x>u$, rillwash is dominant.

Solving for constant downcutting as before,

$$\Lambda = \frac{Tx}{\left[1+\left(\frac{x}{u}\right)^2\right]} \tag{8}$$

In this case, gradient increases for $0<x<u$, and gradient decreases for $x>u$. In other words the constant downcutting form is convex where rainsplash is the dominant process and concave where rillwash is the dominant process (Figure 152).

This relationship is not, however, universal, and if downcutting decreases downslope, the concavity expands slightly into the rainsplash-dominated zone. A simple representation of Davisian decline can be made by assuming, instead of constant downcutting, a rate of downcutting which is directly proportional to height above the basal point. With this assumption, the whole hillslope must eventually erode to a flat uniform base-level plain, and equation (3) is replaced by:

$$-\frac{\partial z}{\partial t} = \frac{\partial C}{\partial x} = \alpha z \tag{9}$$

for an appropriate constant α. Although there is not always a simple analytical solution to this equation, the difference between this Davisian hillslope and the constant downcutting form is quite clear from Figure 153. In each case, the divide convexity begins the same, but, for the declining form, the rate of increase in gradient is less than for the constant downcutting form, and the concavity, where rillwash is present, is broader. It can be shown that the shape of both the declining and constant downcutting forms depends primarily on the combination of processes operating, and the effect of the initial form is progressively obliterated over time. This means that physical remains of former eroded land surfaces (erosion surfaces and terraces) only survive for a limited period of time before they disappear from the landscape, usually surviving longest in flat areas and along divides, where denudation is initially least.

Where travel distances are long, then removal of material is approximately *weathering-limited*. In this case slope development follows equation (2) above, with the final term negligible (because h is large), giving, when combined with equation(1):

$$-\frac{\partial z}{\partial t} = \frac{dS}{dx} = D \tag{10}$$

For the case of mass movements, using equation (3) and writing Λ as $-\frac{\partial z}{\partial x}$:

$$\frac{\partial z}{\partial t} = (\Lambda - \Lambda_T)\frac{\partial z}{\partial x} \tag{11}$$

The solution to this equation shows lateral retreat of steep slopes at a horizontal rate of $(\Lambda - \Lambda_T)$. This evolution of the landscape essentially describes a parallel retreat of the landscape

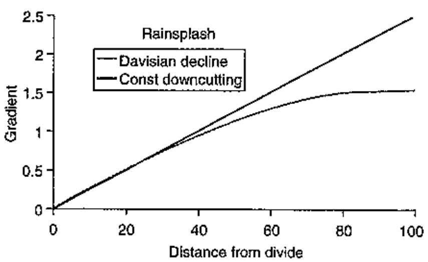


Figure 152 Slope evolution where rainsplash is dominant

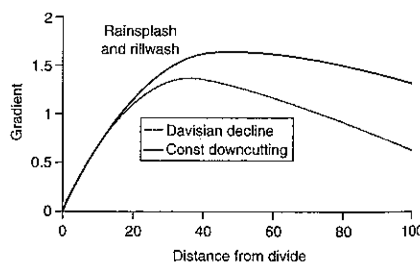


Figure 153 Slope evolution with rainsplash and rillwash

in a way which is close to the conceptual models of Penck and King. Although there is no space to fully develop this argument here, there is a strong association between *transport-limited* removal and Davisian decline of slopes, and between *weathering-limited* removal and lateral retreat of landforms. Because this distinction is closely linked to the presence or absence of a regolith deep enough to allow transport processes to operate at their capacity rates, there is also a general association between these two sets of conditions and both climate and tectonics. There is a link to climate because, in dry climates, there is little water flowing through the soil, consequently little leaching and slow bedrock weathering. Most weathering products are therefore removed by erosional processes before fine-grained and deep soils can develop. On the contrary, humid climates allow more rapid weathering, converting parent materials to fine-textured soils before it is eroded. Steep slopes increase erosion processes, but have much less effect on weathering rates, so that soil is thinner and stonier than on gentle slopes. Active tectonic uplift also plays an important role, by creating and maintaining steep slopes.

From this discussion it may be seen that the various qualitative conceptual models have been significantly shaped by the experiences of their authors. Davis, working in the humid-temperate areas of north-east USA, focused on transport-limited processes, and slope decline towards eventual peneplains. Penck and King, working in more tectonically active and more arid areas, saw the lateral retreat of steep slopes, generally with shallow regolith. It is clear, however, that, although many details are still poorly understood, a single set of principles can be widely applied, and takes in the early conceptual models as special cases.

Figure 154 provides cartoons of 'typical' slope evolution from a plateau with a steeply incised stream, but without further tectonic uplift as material is removed by the basal stream. It can be seen that elements of both transport and weathering-limited removal occur in both settings while slopes are steep, but that weathering-limited elements survive much longer under semi-arid conditions.

Although most of the features of a hillside may be described in a slope profile, the whole scale of the landscape is a problem which can only be addressed in three dimensions, defining the typical length of a single slope, and the DRAINAGE DENSITY of the landscape, which are related by the relationship:

$$\text{Mean slope length} = 1/(2\ \text{DD}) \tag{12}$$

Following the ideas developed by Smith and Bretherton (1972), it is argued that, where sediment transport processes increase more than linearly with catchment area, any small irregularity in the landscape will tend to grow with positive feedback until it develops into a valley. The threshold at which this occurs determines the drainage density of the landscape. A 3-D *landscape model* is able to demonstrate this behaviour. Near the divides, any small irregularities in the landscape become smoother over time, whereas downslope, some hollows grow into valleys. The form of the sediment transport equations, such as equations (3) and (4) above, determine the threshold of this valley instability. For example, if the combination of rainsplash and wash is put in the form (note that this is a different form from equation (7) above):

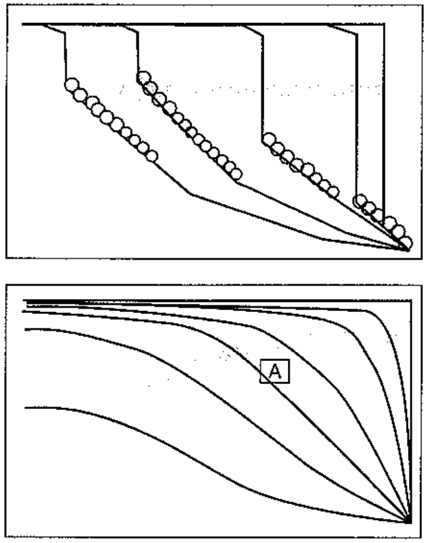
$$S = C = k\Lambda\left[1+\Lambda\left(\frac{x}{u}\right)^2\right] \tag{13}$$

then it can be shown that the critical distance for hollow enlargement is:

$$x = u / \sqrt{\Lambda} \tag{14}$$

Table 44 Conditions associated with transport-limited and weathering-limited removal

	Transport-limited	*Weathering-limited*
Regolith	Deep enough to allow transport processes to operate	Shallow and generally stony
Climate	Humid temperate	Semi-arid
Gradient	Gentle	Steep
Tectonics	Inactive	Active
Dominant erosion processes	Creep, rainsplash, rillwash	Mass movements
Ratio of weathering to erosion	High	Low



Semi-arid slope evolution
Upper plateau influenced by rainsplash etc. Steep escarpment (often defined by lithology) and boulder slope influenced by mass movements. This is the only section undergoing parallel retreat. Lower concavity modified by rillwash and stream incision.

Humid-temperate slope evolution
Upper plateau influenced by creep or solifluction. Until A, lower slope dominated by mass movements, at reducing rate as slope towards a landslide-stable gradient. After A, slope dominated by creep and rillwash, declining towards base-level peneplain.

Figure 154 Slope evolution from a plateau with a steeply incised stream under semi-arid and humid-temperate conditions

This expression suggests that drainage density should be greater in steeper areas, and this forecast is supported by empirical evidence, particularly Dietrich and co-worker's data from California (Dietrich and Dunne 1993). Thus the spacing of streams, and so the whole scale of the landscape, is also set by the balance between slope and stream processes, and valleys occur where the processes driven by water flow begin to predominate over processes driven mainly by gradient, such as creep or rainsplash. This view of the landscape indicates not only that drainage density varies over the landscape in relation to steepness, but that it evolves through time as relief changes through erosional lowering or tectonic uplift.

References

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.

Davis, W.M. (1909) *Geographical Essays*, Boston: Ginn; (1954), New York: Dover.

Dietrich, W.E. and Dunne, T. (1993) The Channel Head, in K. Beven and M.J. Kirkby (eds) *Channel Network Hydrology*, 175–219, Chichester: Wiley.

Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, Washington, DC: US Geological and Geographical Survey.

Hack, J.T. (1960) Interpretation of erosional topography in humid temperate regions, *American Journal of Science* 258A, 80–97.

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

Kirkby, M.J. (1971) Hillslope process-response models based on the continuity equation, *Institute of British Geographers Special Publication* 3.

Penck, W. (1953) *Morphological Analysis of Landforms*, trans. by H. Czech and K.C. Boswell, London: Macmillan.

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.

MIKE KIRKBY

SLOPE STABILITY

Slope stability and its corollary slope instability, are defined as the propensity for a slope to undergo morphologically disruptive processes, especially landsliding (Plate 125) (see LANDSLIDE). Slow distributed forms of MASS MOVEMENT such as soil creep are generally not considered sufficiently disruptive to be included in this definition. From a hazard and engineering perspective, assessments of slope stability are focused on periods ranging from days to decades. However, slope stability may also be treated as a component of landform evolution and therefore its significance can only be judged by taking into account much longer periods of time.

In every slope, there are stresses which tend to promote downslope movement of material (shear stress) and opposing stresses which tend to resist movement (shear strength). In order to assess the degree of stability, these stresses can be calculated for a failure surface within the slope and compared to provide a FACTOR OF SAFETY (defined as the ratio of shear strength:shear stress). In a static slope, shear strength exceeds shear stress and the factor of safety is greater than 1.0 whereas for slopes on the point of movement shear strength is just balanced by shear stress and the factor of safety is 1.0.

While engineering codes of practice may specify a particular factor of safety for completed earthworks, it is not the most meaningful representation of slope stability. Two slopes that have the same factor of safety but large absolute differences in excess strength (i.e. strength minus shear stress) can be used to illustrate the limitations of the factor of safety. For example, the strength to stress ratios, in unspecified stress units for a slope (A) of 400/200 and for a slope (B) of 200/100 both yield a factor of safety of 2.0. However, slope (A) has an excess strength of 200 units while slope (B) has an excess strength of only 100 units. As excess strength is the quantity that must be entirely reduced (by reduction in strength or increase in shear stress) in order to produce failure, it represents the 'margin of stability' or inherent resistance to failure. Instability, however, is determined not only by the margin of stability of the existing slope but also by the magnitude of (external) destabilizing forces which may affect the slope to reduce that margin.

Plate 125 Actively unstable slopes, subject to deep-seated earthflows, Poverty Bay, New Zealand. Photo: Ministry of Works, New Zealand

Slopes can therefore be viewed as existing at various points along a stability spectrum ranging from high margins of stability with low probabilities of failure at one end to actively failing slopes, with no margin of stability, at the other. It is useful to define three theoretical stability states along this spectrum based on the ability of dynamic external forces to produce failure. First is the 'stable state', defined as slopes with a margin of stability which is sufficiently high to withstand the action of all dynamic destabilizing forces likely to be imposed under the current environmental/geomorphic regime. Second is the 'marginally stable state', represented by static slopes, not currently undergoing failure, but susceptible to failure at any time that dynamic external forces exceed a certain threshold. Third is the 'actively unstable state', represented by slopes with a margin of stability close to zero and which undergo continuous or intermittent movement. The margin of stability possessed by a slope is a measure of its landslide susceptibility and, together with the frequency and magnitude of dynamic destabilizing factors, provides a measure of probability of failure. In turn, the probability of failure together with its magnitude provides a measure of landslide hazard.

The concept of three stability states offers a useful framework for understanding the causes and development of instability. In this context four groups of destabilizing factors can be identified on the basis of function (Figure 155).

1 *Preconditions* (predisposing factors) are static, inherent factors which not only influence the margin of stability but more importantly in this context act as catalysts to allow other dynamic destabilizing factors to operate more

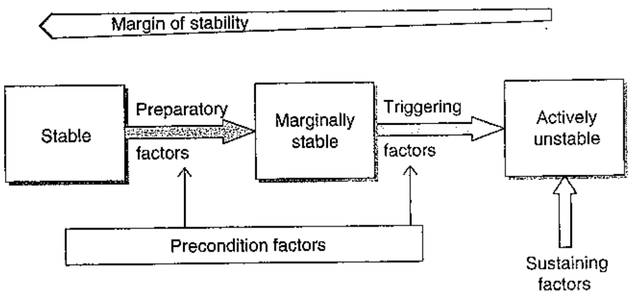


Figure 155 Stability states and destabilizing factors

effectively. For example, slope materials that lose strength more readily than others in the presence of water predispose the slope to failure during a rainstorm; or a particular orientation of rock structure may enhance the destabilizing effects of undercutting.

2 *Preparatory factors* are dynamic factors that by definition decrease the margin of stability in a slope over time without actually initiating movement. Hence, facilitated by preconditions, they are responsible for shifting a slope from a 'stable' to a 'marginally stable' state. Some factors, such as reduction in strength by weathering, climate change and tectonic uplift, operate over long periods of geomorphic time whereas others may be effective in shorter time periods e.g. slope oversteepening by erosional activity, deforestation, or slope disturbance by human activity.

3 *Triggering factors* are those factors which initiate movement, i.e. shift the slope from a 'marginally stable' to an 'actively unstable' state. The most common triggering factors are intense rainstorms, seismic shaking and slope undercutting.

4 *Sustaining* factors are those that dictate the behaviour of an 'actively unstable' slope e.g. duration, rate, and form, of movement.

As most slopes are stable for most of the time, the onset of slope instability from a geomorphic perspective represents an effective hillslope response to a destabilizing change in the boundary conditions, enabling a rapid adjustment and eventual return (or tendency toward) more persistent landscape forms. Thus, given sufficient time, the process of landsliding tends to stabilize a slope by reducing slope angle, height or weight, or by removing susceptible material.

The concept of slope instability may usefully be broadened to include any significant adjustment in processes or forms that tend to change the equilibrium conditions on a slope. For example, a slope where mature soils are being depleted by the onset of gullying can be considered to be undergoing a phase of slope instability. Whether that adjustment is viewed as a perturbation of a system in dynamic equilibrium or a change to another equilibrium state depends on the timescale being considered.

Further reading

Crozier, M.J. (1989) *Landslides: Causes, Consequences, and Environment*, London: Routledge.
Selby, M.J. (1993) *Hillslope Materials and Processes*, Oxford: Oxford University Press.

MICHAEL J. CROZIER

SLOPEWASH

The term 'slopewash' is not well defined in geomorphological literature. It has been used interchangeably with terms such as 'surface wash', 'rainwash', 'unconcentrated wash' and 'rillwash', now largely replaced by more precise terms such as 'overland flow', 'sheet wash', 'rainflow', 'rain-impacted flow', and 'rilling' or 'rill erosion'. The process was first described by McGee (1897), and defined by Bryan (1922: 29) as 'the water from rain, after it has fallen on the surface of the ground and before it has concentrated into definite streams'. Surface wash and rainwash were extensively invoked in early geomorphological literature as processes responsible for concave hillslope profiles, for hillslope profiles with a marked concave 'break', and for the formation of rock-cut pediments in southern Africa (King 1949). However, few field studies (e.g. Schumm 1956) attempted to measure the process, and it was not until Emmett's (1970) study of the hydraulics of overland flow on hillslopes that the complexity of the process was recognized. Subsequent studies on agricultural soils have identified the precise interacting processes involved, which include rainflow, sheet wash and rill erosion. The processes combined in slopewash are frequent on disturbed agricultural soils but their significance on natural hillslopes is less clear. The extensive thin sheets of water described by McGee are rare, even in dry regions where conditions are most favourable. In all areas, overland flow occurs most frequently on bare, relatively impermeable rock surfaces. Depending on the rock, such surfaces will usually yield some fine-grained material which can be transported by shallow flows, but the dominant transport in most cases is probably solution. Overland flow generated where rainfall exceeds surface infiltration capacity (Hortonian overland flow) occurs quite frequently on regolith-covered hillslopes, but is usually localized, reflecting patchy vegetation cover, microtopography, variations in rainfall intensity and marked variations in surface infiltration capacity. This produces a complex, varied interaction of processes along the hillslope, rather than the orderly transition from rainsplash to sheet wash to rill erosion envisaged in earlier literature. Instead, on most hillslopes patches of erosion, erosional lag deposits and sedimentation are interspersed in complex patterns. Often these are random and irregular, but on dryland hillslopes, the patterns are often regular, caused by the shrink–swell characteristics of smectite clays, the moisture requirements of shrubby plants, or selective transportation of material by pulsatory flows. These features may also occur where flow is not Hortonian, but generated as saturation excess, as return flow or as seepage. Such flows are more predictable and usually more uniform across the hillslope and therefore may more frequently cause profile concavity.

References

Bryan, K. (1922) Erosion and sedimentation in the Papago Country, Arizona, *United States Geological Survey Bulletin*, 730B, 19–20.
Emmett, W.W. (1970) The hydraulics of overland flow on hillslopes, *United States Geological Survey Professional Paper* 730B.
King, L.C. (1949) The pediment landform: some current problems, *Geological Magazine* 86, 245–250.
McGee, W.J. (1897) Sheetflood erosion, *Geological Society of America Bulletin* 8, 87–112.
Schumm, S.A. (1956) The role of creep and rainwash on the retreat of badland slopes, *American Journal of Science* 254, 693–706.

RORKE BRYAN

SLUSHFLOW

Slushflows, also called slush avalanches, are water-saturated snow masses flowing principally along a first-order stream channel (Larocque *et al.* 2001). Their formation is associated with an increase in the water content of a snowpack through rainfall, or through rapid snowmelt or through a combination of both. At a critical point instability occurs and snow mass is released. They are widespread in artic, subarctic and alpine environments but may, unlike snow avalanches (see AVALANCHE, SNOW), be initiated on relatively low angle slopes (Nyberg 1989). They are capable of transporting a high debris load over long distances, and also of causing substantial abrasion and erosion. The material they deposit can form such features as boulder tongues. They can also cause exceptional rates of glacial ablation (Smart *et al.* 2000) and pose a hazard to engineering structures.

References

Larocque, S.J., Hétu, B. and Filion, L. (2001) Geomorphic and dendroecological impacts of slushflows in Central Gaspé Peninsula (Québec, Canada), *Geografiska Annaler* 83A, 191–201.
Nyberg, R. (1989) Observations of slushflows and their geomorphological effects in the Swedish Mountain area, *Geografiska Annaler* 71A, 185–198.
Smart, C.C., Owens, I.F., Lawson, W. and Morris, A.L. (2000) Exceptional ablation arising from rainfall-induced slushflows: Brewster Glacier, New Zealand, *Hydrological Processes* 14, 1,045–1,052.

A.S. GOUDIE

SOIL CONSERVATION

Society can try to minimize SOIL EROSION by soil conservation practices, which may be 'active' or 'passive'. Active soil conservation includes positive actions to decrease erosion rates, such as terracing or contour farming. 'Passive' conservation is avoiding actions, such as ploughing on steep slopes or overgrazing erodible soil. Passive conservation is just as valid and often much cheaper than active conservation. We also need to distinguish between soil conservation and sediment control. Soil conservation is taking action to prevent erosion and keep soil *in situ*. Sediment control deals with soil which has already been eroded and transported, keeping it within fields or removing it from water courses.

Conservation strategies can be complex and varied. However, Morgan (1995) proposed that soil is conserved by decreasing EROSIVITY, decreasing ERODIBILITY, improving vegetation protection, or any combination of these. Ancient features, such as the terraced fields of South-East Asia, show that soil conservation has long been employed. However, soil conservation technologies were considerably improved by the US Soil Conservation Service.

The US Midwest suffered severe soil erosion in the 1930s or 'dirty thirties'. By this time, many of the organic soils had been cultivated for over eighty years, so soil organic contents were falling, thus increasing soil erodibility. Many erodible soils were brought into cultivation due to increased grain prices during the First World War. Then, in the 1930s, there was a severe drought. This combination of circumstances led to severe erosion, especially wind erosion, in the Great Plains States, which became labelled 'The Dust Bowl'. The social upheaval associated with these events was graphically portrayed in the novel *The Grapes of Wrath* (Steinbeck 1939), which tells the tragic struggle of a family from Oklahoma.

In response to these problems, an agricultural engineer, Hugh Hammond Bennett, led a campaign to promote soil conservation. Bennett was almost evangelical in his campaigning among farmers and politicians. During one presentation to the US Congress, a dust storm blew into Washington, DC. Bennett declared, pointing out the window, 'there, gentlemen, goes the State of Oklahoma!'. Congress then allocated funding for the foundation of the US Soil Conservation Service in 1934.

Various soil conservation techniques have been proposed and conservation projects can adopt one or a combination of techniques. The choice depends on many factors, including environmental (climate, topography and soil type) conditions and social, economic and political circumstances.

Windbreaks can be very effective in decreasing wind velocity and thus erosivity. They are usually aligned perpendicular to the direction of the most frequent erosive winds and are effective for 10–20 times their height downwind. Dense windbreaks greatly decrease wind velocity, but velocities soon increase in their lee. Better effects are achieved with more permeable windbreaks; velocity is not decreased so much, but their effectiveness in the lee is greater. For maximum benefit, windbreaks with a porosity of about 50 per cent and a mix of heights are recommended.

Hedgerows are effective windbreaks and their large-scale removal from the British countryside since the 1940s has contributed to increased wind erosion of arable soils. Hedgerows also protect against water erosion, by dividing slopes into shorter sections. Their removal allowed erosive runoff to operate over effectively longer slopes, thus increasing erosion risk. The restoration of all hedgerows is not feasible, as modern farm machinery cannot operate efficiently on fields as small as those which dominated the British countryside before the Second World War. However, it is important to identify 'key hedgerows', that protect areas exposed to predominant wind directions and convex slope sections susceptible to water erosion. Their retention or establishment should separate large catchment fields from lower convex–concave slopes. The absence of hedgerows integrates fields into geomorphological systems that are vulnerable to water erosion.

Slope management is a key component of soil conservation strategies. Simply leaving steeper erodible soils with a vegetation cover is a cheap, but effective, form of passive conservation. 'Set-aside', which involves taking areas out of crop production and leaving them with a permanent vegetative cover, was originally a means of decreasing grain surpluses. Carefully targeted on steeper erodible slopes, it could be an effective means of achieving both agricultural and environmental objectives.

Terracing is the most spectacular form of soil conservation and involves dividing slopes into a series of steps, cultivating the flatter sections and protecting the 'riser' with vegetation or masonry walls. In South-East Asia, terraces are extensively used to grow rice. To retain water, small earth walls or 'bunds' are built on the lower sides of terraces. However, terracing poses several problems. First, the risers take up about 10 per cent of land, though this is usually compensated by increased crop yield associated with increased soil water storage. Second, construction and maintenance are costly in terms of human resources; most of the world's areas of extensive terraces were constructed over many generations. Third, it is often difficult to operate machines efficiently on terraces.

Controlled colluviation is particularly applicable in semi-arid countries with a distinct rainy season. Lines of stones are laid out along the contour. When the seasonal rains arrive, soil is eroded and sediment deposited on the upslope side of the stones. Over a few rainy seasons, a fine silty moisture-retentive colluvial soil accumulates and, in semi-arid climates where soils tend to become saline, seasonal flushing with water can desalinize the COLLUVIUM, making it relatively fertile and suitable for crops. The technique is simple, cheap and thus affordable in poorer countries. Check dams operate in a similar way. Walls are constructed across gullies, sediment collects behind them and is periodically removed. Additionally, obstacles (such as straw bales and willow fences) are often placed in gullies to impede runoff and thus encourage sedimentation.

Contour farming involves orientating agricultural operations along the contour, rather than up-and-down slope and is particularly applicable on gentle uniform slopes used for mechanized agriculture (e.g. the Prairies and the Steppes). Complex slopes tend to limit the applicability of contour cultivation in northern Europe. However, slopes must not be too steep (>15°), as water can accumulate in furrows and eventually breach the ridges between them, causing even higher soil erosion rates than the more common up-and-down slope cultivation. Moreover, farm machinery cannot operate safely or efficiently along the contours on steep slopes.

In strip cropping systems, alternate strips of land are arranged perpendicular to the relevant erosive agent, wind or water. The crops themselves protect the soil, braking the velocity of the erosive agent and trapping sediments. Temporary grassland (leys) can form part of a strip cropping system and also increase soil organic matter, lowering soil erodibility. Usually, strips are 15–45 m wide, becoming wider as erosion risk increases.

Rotation is a well-established agronomic technique, by which different crops are grown in an established sequence. A temporary grass ley is usually an integral component of rotational systems, allowing natural recovery. Twentieth-century development and mass production of chemical fertilizers allowed continuous arable production, without temporary leys. Fertilizers provided ample supplies of macronutrients needed for crop production but, on many soils, allowed soil organic contents to decrease, so that erodibility increased. The increased incidence of erosion on arable soils in much of North America and Western Europe has been attributed to long-continued arable cultivation.

Addition of organic matter can decrease soil erodibility. The most common is farmyard manure (FYM) and there are many commercial organic manures. 'Green manures' are crops such as clover and mustard which grow quickly, producing much biomass, but rapidly decompose to increase the soil organic matter. In developing countries, human waste is used. This is usually transported and applied at night and is known as 'night soil'.

Mulching is the use of vegetative or other material on the soil surface, to simulate the protective effects of vegetation. Often, residues from the previous crop are applied, such as straw on wheat fields. Many studies have shown mulching is very effective, especially in the tropics and subtropics. However, in temperate environments, mulches can insulate the soil and prevent it warming in spring.

Hydromulching is particularly applicable to engineered slopes, such as road cuttings and construction sites. Mulch consists of a mixture of materials, such as fibre, straw, paper and shredded wood and is sprayed with seed, protecting the soil surface while seeds germinate and establish a protective vegetation cover. Hydromulching is expensive and used only in high value projects.

Geotextiles are cloths used to protect soil surfaces. Usually, they consist of biodegradable material, such as jute, have a coarse mesh and last for a short time, usually less than two years. This is long enough for seed mixtures to establish protective plant communities. There are also non-biodegradable geotextiles, which are used for permanent stabilization of, for example, channel banks.

Compaction affects the hydrological and thus erosional behaviour of soil, decreasing the size and interconnectivity of soil pores and impeding infiltration, so that runoff and erosion are increased. Increasing use of heavy farm machinery is exacerbating compaction problems. Compaction by animals also poses problems, as they produce fairly small compact hoof imprints. On wet soils, this weakens soil structure, and gives soil a 'puddled' appearance, a process known as poaching.

There are many methods of minimizing soil compaction. Trafficking wet soil should be avoided, though a shallow tillage tool mounted behind the tractor wheel can break up compacted soil. Tramlines, along which all passes for agricultural operations are made, limit compaction to narrow zones. Even within tramlines, it is possible to diminish compaction by using larger tyres or low-ground pressure vehicles (LGPV) to spread the load. Ploughing in very wet conditions and/or with blunt plough shares can compact subsoils. This 'plough pan' must then be disrupted by 'subsoiling', using a deep blade with a 'shoe-like' structure at the end, mounted behind a powerful tractor.

Several crop production methods (known variously as crop residue management systems, no-tillage, minimum cultivation, conservation tillage or direct drilling) minimize compaction, runoff and loss of organic matter by avoiding ploughing and preserving a cover of crop residues. The various terms partly reflect the diversity of systems in use. Residues from the previous crop are left on the soil surface, to simulate the protective effects of vegetation. Further crops are then sown into the residue with minimum disturbance and eventually provide a protective vegetative cover. These systems allow crop production on steeper slopes without increasing erosion risk. Increased soil organic matter also improves moisture retention and encourages earthworms, which increase infiltration rates. There are disadvantages, as lack of tillage can allow weed infestation, particularly by perennial grasses. As compaction can increase without tillage, especially on weakly structured silty soils with little organic matter, a crucial factor for success is soil faunal activity, especially earthworm activity. A rich earthworm population can effectively till the soil on behalf of the farmer. This is why conservation tillage systems became popular on organic-rich soils in both North and South America. There is a gradation from full 'no-till' to traditional ploughing. The selected method depends on the various advantages and disadvantages. A minimum cultivation system, with occasional full ploughing, can be an acceptable balance.

Chemical soil conditioners can decrease soil erodibility, binding particles together, improving aggregation and increasing infiltration rates. Conditioners, such as 'Krilium', 'Flotal' and 'Glotal', were developed in the 1950s and 1960s, and were followed by many ionic and non-ionic conditioners. Field and laboratory experiments showed they were effective, but high cost restricted their use to high value crops or specialist engineering applications (e.g. stabilizing road cuttings, engineered slopes, oil-well heads, temporary helipads and airfields).

Successful soil conservation is not simply an engineering problem, but is a complex amalgam of agronomic, social, economic and political considerations. It is essentially a team effort, requiring the participation of national, provincial and municipal government, farmers, scientists, extension workers and agricultural advisers. Also, it is essential that an effective dialogue develops between soil conservationists and local people, as their involvement, support and participatory agreement are crucial.

References

Morgan, R.P.C. (1995) *Soil Erosion and Conservation*, 2nd edition, London: Longman.

Steinbeck, J. (1939) *The Grapes of Wrath*, New York: Milestone Editions.

MICHAEL A. FULLEN AND JOHN A. CATT

SOIL CREEP

Slow MASS MOVEMENT processes are generally grouped together under the term soil creep (Kirkby 1967). A distinction is made between continuous creep, which is driven directly by downslope shear stresses, and seasonal creep which is the downslope component of movements which are either randomly directed, or primarily perpendicular to the soil surface.

Continuous creep may extend to depths of 5 m or more, and is driven by gravity shear stresses against the frictional and cohesive strength of the soil. The relationship between stress and strain for soils usually shows three zones. At low stress there is no movement; at high stress there is a more or less linear deformation, at a rate proportional to stress above an apparent failure threshold; at stresses below this threshold there is slow and non-linear increase in strain which bridges between these two behaviours (Figure 156). Continuous creep occurs in this non-linear transitional zone. Clearly continuous creep can occur over only a narrow range of stress–strain conditions, and is, in most cases, associated with the larger mass movements which occur when stresses appreciably exceed the threshold. In many soil materials, the creep itself changes the stress–strain relationship as the soil is reworked by movement.

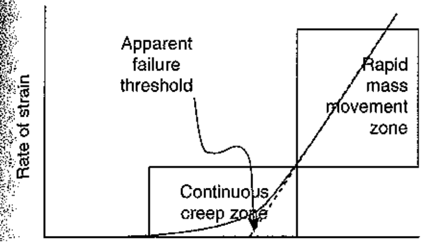


Figure 156 Stress–strain relationship in continuous creep

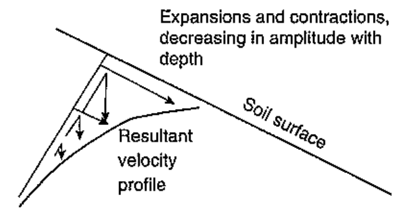


Figure 157 Zigzag seasonal creep movement

Seasonal soil creep is driven by a number of processes, some of them biogenic and others driven by the climate. Biogenic movements are more or less random in direction, and include wedging by plant roots, movement of soil by burrowing animals such as gophers, earthworms and termites. Random movements lead to a net migration of material from zones of high soil concentration to zones of lower concentration, and may be considered as a form of diffusion.

The net direction of this random migration is towards the free upper surface of the soil, and it is eventually balanced by resettlement under gravity. The two main climate drivers are WETTING AND DRYING WEATHERING cycles and FREEZE–THAW CYCLES. These cycles lead to an upward expansion or heave during wetting or freezing, followed by a downward settlement during drying or thawing. On a slope these expansions are at right angles to the slope surface, while settlement tends to occur more nearly along a vertical (Figure 157). Thus both random movements and heaves driven by climate produce a net movement which consists of zigzags which move material slowly downslope. To a first approximation the rate of this movement should be proportional to the slope gradient, but this theoretical inference has never been completely validated, because of the large observed variability in creep rates over small areas. All these seasonal soil creep processes are commonly referred to as *diffusive processes*, and have in common a linear dependence of transport rate on slope gradient, and little or no dependence on distance from divide or catchment area because they do not depend on the flow of water.

This balance between outward diffusion and settlement under gravity not only moves material downslope but is also one of the main processes responsible for the observed increase in pore space towards the surface of uncultivated soils.

It can be seen that the rate of seasonal soil creep is limited by the depth and amplitude of these random and heave movements, which are always small. Thus there can never be a transition from seasonal soil creep to larger and more rapid mass movement, and it rarely extends to depths where the movement reworks material along potential mass movement failure surfaces. In different areas, the dominant processes reported to drive seasonal soil creep can be almost any of those listed above, according to the faunal and climatic activity in the soil. Seasonal soil creep rarely penetrates to depths of more than 30 cm, and surface translational movements reported generally lie in the range of 2–5 cm per year. If movement is proportional to gradient, then a widely quoted diffusivity for soil creep is $10^{-3}\,m^2yr^{-1}$. This means that on a 10 per cent slope, the actual transport will be $10\% \times 10^{-3}\,m^2yr^{-1}$, or 1 $cm^2\,yr^{-1}$. These values are low, of the same order of magnitude as those quoted for rainsplash (see RAINDROP IMPACT, SPLASH AND WASH), whereas the highest values reported for SOLIFLUCTION are 10–100 times greater. Both these processes are commonly included with soil creep in the category of diffusive processes, at least to a first approximation. Another important anthropogenic diffusive process is termed *tillage erosion*, in which ploughing and other forms of cultivation produce considerable net downslope movement of material, even if successive passes of the plough are along the contour, and alternately turn the soil up- and down-slope. Rates of tillage erosion are currently 10–1,000 times greater than for soil creep or rainsplash, and have increased greatly with the more widespread use of heavy farm machinery.

G.K. Gilbert (1909) recognized the importance of diffusive processes in creating convex hilltops, which are found on almost all soil-covered slopes. The convex area is usually relatively smooth, and may give way to an area where rills or larger channels begin, or may lead directly into an area where slopes become more uniform in gradient. The width of this convexity varies widely, primarily with climate and lithology, and generally lies in the range 1–1,000 m. Typically narrow convexities are associated with high DRAINAGE DENSITY and broad convexities with low drainage density. The relationship between diffusive processes and hillslope convexity is based on the assumption that transport by diffusive processes increases with gradient, and little or not at all with distance from the divide (or catchment area). If the area around the divide is eroding in the long term, then the volume of material which must be transported to remove the erosion products necessarily also increases continuously away from the divide. The diffusive processes can, by definition, only transport this increase in material through an increase in gradient. If the REGOLITH is deep enough not to constrain movement of material, it necessarily follows that gradient must also increase continuously from the divide, i.e. that the slope profile is convex.

This argument makes very few assumptions, so that the conclusion – that divides are convex – has great generality, and requires only that diffusive processes are dominant close to the divide. It is also fair to invert the argument, since this is a necessary and sufficient condition, and state that divide convexity demonstrates the dominance of diffusive processes close to the divide, and indeed roughly outlines the zone in which they are dominant. In the simplest realistic case, denudation of the area around the divide may be assumed to be constant, at rate T. At distance x from a ridge crest, an amount $T \cdot x$ must be exported per unit length of ridge crest to carry away the eroded material. If the rate of transport by creep processes $= K \cdot s$, where s is slope gradient and K is the rate constant, then to balance the erosion against the transport rate, $T \cdot x = K \cdot s$ or the slope must evolve until the gradient $s = T/K \cdot x$. In other words the gradient must eventually increase linearly with distance from the ridge crest.

It is useful to consider what happens if the assumptions of this argument are not met. First, on a long valley side, there may be deposition at the base of the slope, so the assumption of uniform erosion breaks down, and diffusive processes may be associated with concavity at the slope base. Second, if the regolith is thin over a resistant rock layer, then the K value may drop where the soil is thin, and rise again below the resistant layer, so that there is exaggerated convexity over the rock band, and a short concavity below, as the K value rises to its previous value.

At high gradients, there is some departure from linearity for creep and other diffusive processes. As the gradient approaches a threshold of stability (say 35° or 70 per cent), then material will roll downslope with minimal disturbance, and the rate of sediment transport increases rapidly with increasing gradient. Instead of a linear dependence on gradient, s, sediment transport can be estimated as:

$$\frac{K \cdot s}{(1 - s/s_*)^m}$$

where s_* is the threshold gradient, and the exponent $m = 1 - 2$. Applying the same argument for the slope profile, non-linear diffusion generates a summit convexity which, if there is sufficient relief, straightens out downwards towards a uniform slope at gradient s_*.

In the summit area where diffusive processes are dominant, SLOPE EVOLUTION creates a stable regime in which any minor irregularities in the hillslope surface tend to be obliterated rather than enlarged. Thus the divide area is generally smooth at a scale larger than that of microtopography, and any initial irregularities are progressively removed by the diffusive processes. This broad-scale smooth convex form may sometimes be observed on slopes where there is no soil, particularly on limestones in semi-arid and tropical climates. It may generally be inferred that the convexity was developed under a continuous soil cover, which has since been removed by erosion or washed into joints enlarged by solution.

Seasonal soil creep normally extends to depths of 30–50 cm, and the transport processes act at their full capacity provided that the regolith exceeds this depth. To maintain the rate of soil creep, WEATHERING processes must replace the soil lost by denudation, which is likely to be at about $100\,\mu m\,yr^{-1}$ for typical rates and slope forms, increasing with convexity and with the diffusive rate K. Weathering in this context refers to the physical breakdown of rock or *saprolite* (regolith which has been chemically altered *in situ*, retaining the original parent material structure), usually by biogenic or cryogenic processes. Dating of the cosmogenic (see COSMOGENIC DATING) isotope Beryllium-10, from the base of soils in California where bioturbation was by rodents (pocket gophers), Heimsath *et al.* (2001) showed a rate of weathering which decayed exponentially with soil depth, of approximately $280 \exp(-z/30)\ \mu m\,yr^{-1}$ at the base of a soil of depth z (cm). To maintain equilibrium between this rate of biogenic weathering and a denudation rate of $100\,\mu m\,yr^{-1}$, the soil should therefore be approximately 31 cm deep. This equilibrium analysis correctly forecasts, for example, that, in comparing the gentle convexities typical of temperate landscapes with the much sharper and narrower convexities of semi-arid badlands, generally thinner regolith is developed on the semi-arid divides.

In summary, soil creep is defined as any slow mass movement. Continuous creep generally occurs in soils which are also subject to rapid mass movements, at shear stresses just below their failure threshold. Seasonal soil creep occurs in a wide range of soils, due to biogenic activity, freeze–thaw and wetting–drying, and the rate of sediment transport is usually assumed to increase linearly with slope gradient. Together with other diffusive processes, it is responsible for developing bulk density profiles and summit convexities. In equilibrium with weathering processes, seasonal creep also determines the depth of the regolith on convexities.

References

Gilbert, G.K. (1909) The convexity of hilltops, *Journal of Geology* 17, 344–350.

Heimsath, A.M., Dietrich, W.E., Nishizumi, K. and Finkel, R.C. (2001) Stochastic processes of soil production and transport: erosion rates, topographic variation and cosmogenic nuclides in the Oregon Coast range, *Earth Surface Processes and Landforms* 26, 531–552.

Kirkby, M.J. (1967) Measurement and theory of soil creep, *Journal of Geology* 75, 359–378.

MIKE KIRKBY

SOIL EROSION

Water and wind can erode, transport and eventually redeposit soils. The initial impact of raindrops can break soil aggregates into primary particles, by the translation of kinetic energy from

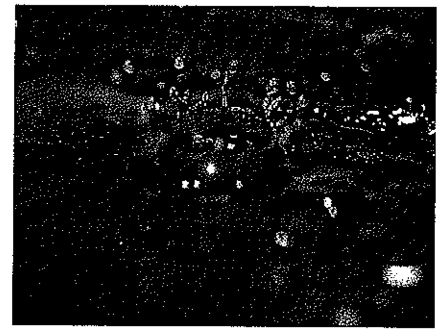


Plate 126 A simulated raindrop hitting a moist sandy surface. The kinetic energy of the impacting raindrop causes the formation of an impact crater, around which the water rebounds as a corona. The process occurs very quickly – this photograph was shot at 1:2,000 of a second

the drops to the soil aggregates, a process known as SLAKING (Plate 126). Due to the influence of gravity, more soil particles are splashed downslope than upslope, and detached particles are splashed further downslope. The cumulative effect is a net downslope transfer of soil particles, known as splash erosion.

When rainfall intensity exceeds soil infiltration capacity, runoff occurs. Initially, a sheet of water of fairly uniform thickness flowing over the surface causes quite uniform sheet erosion (see SHEET EROSION, SHEET FLOW, SHEET WASH). However, this state is unstable, as flowing water concentrates in surface depressions and incises into the soil. Where these channels are shallow, the process is RILL erosion. However, if rill erosion continues gullies develop.

The distinction between rill and GULLY erosion is problematic. Early guidelines from the US Soil Conservation Service stated a gully is too wide for a prairie dog to jump across! Later definitions stated a gully would need to be mechanically infilled for agricultural activities to proceed. Others argue that rills are incised into the topsoil (the A horizon), whereas gullies incise into the subsoil or parent material (the B or C horizons).

Eroded soil is eventually redeposited. When this occurs on slopes, the redeposited material is termed COLLUVIUM. It accumulates on concave sections of slopes or against obstructions, such as walls and hedges, as decreasing flow velocities decrease the transport capacity of the runoff water. Sediment can also enter water courses, where it is reworked to form ALLUVIUM. Sediments derived from farmland soils are often rich in agrochemicals, such as phosphate, nitrate and pesticides and can pollute water and damage aquatic ecosystems.

Wind erosion (AEOLIAN PROCESSES) can transport fine sediment considerable distances. For instance, at the Mauna Loa Observatory in Hawaii, the onset of the Chinese spring planting season results in increased dust fallout, even though Hawaii is 5,000 km east of China (Parrington *et al.* 1983). Finer particles, especially silt, are usually transported in suspension above the land surface; sand is usually transported by SALTATION (particles bouncing along the surface).

The severity of soil erosion varies markedly. Langbein and Schumm (1958) proposed a model relating water erosion to rainfall. In very dry climates, there is usually little water erosion, but a slight increase in rainfall often causes a large increase in erosion rates. The occasional rainstorms tend to be very intense convectional storms, which have considerable energy to cause erosion. Also, there is little protection from vegetation, which is sparse in semi-arid environments. In more humid climates, greater vegetation cover protects the soil surface from erosion.

The Langbein and Schumm model has been critically assessed in many parts of the world. Certainly, in the semi-arid tropics, erosion rates are very high. It is estimated that some 6,000 million tonnes of soil per year are washed off the croplands of India and that 1,600 million tonnes per year are transported by the River Ganges to the Bay of Bengal. The Mediterranean environment also has high rates because of the semi-arid climate and long human occupancy of the landscape, which has promoted deforestation (Plate 127).

There are areas where erosion rates should be low, according to the Langbein and Schumm model, but actually are high. This largely reflects human activities, particularly vegetation removal. For instance, destruction of tropical rainforests has greatly increased erosion rates. Rainforest soils contain little organic matter, unlike those of more temperate environments, and removal of vegetation can lead to further losses. Organic matter stabilizes soil structure, so its loss increases erosion risk.

Temperate continental interiors (Prairies of North America, Steppes of Central Asia, Pampas of Argentina) should have a natural vegetation cover of grassland. Under these conditions, the soils are thick, black organic Chernozems. Over the past 100–130 years many of these grasslands have been converted to arable use, particularly for cereal production. The soil organic content has decreased and the soils have become more ERODIBLE. Some 4,000 million tonnes of soil per year have been eroded from the continental USA since the 1930s. This would fill a freight train long enough to encircle the equator twenty-four times!

There is increasing evidence of soil erosion in northern and western Europe. In the Langbein and Schumm model, these areas should experience little erosion. However, the damage to soil structure imposed by increasingly mechanized and intensive agriculture, particularly since the 1940s, is believed to have increased erosion rates. Erosion directly related to cultivation is often termed tillage erosion.

Soil erosion can be described as 'a quiet crisis, one that is not widely perceived. Unlike earthquakes, volcanic eruptions or other natural disasters, this human-made disaster is unfolding gradually' (Brown 1984). Brown estimated that the maximum rate of soil formation is 2 tonnes of soil per hectare of land per year, but that the average soil erosion rate is about 20 $t\,ha^{-1}\,yr^{-1}$. The calculated global excess of soil erosion over formation is 25,730 million tonnes, roughly equivalent to the amount of topsoil in Australia's wheatlands. These estimates are very tentative, but they indicate the scale of the problem.

It is difficult to define what is 'acceptable' as a rate of soil erosion. 'Natural' or 'geologic' erosion is a natural process by which uplands are denuded over geological time. This is different from 'accelerated' erosion, where human activities are increasing erosion rates much above their background levels. The commonest activity is vegetation removal, exposing soils to erosion.

To define 'acceptable' soil erosion, many soil scientists suggest that the soil system should be in a state of dynamic equilibrium; that is, the rate of loss by erosion should not exceed the rate of soil formation. We know very little about rates of soil formation, except that it is a slow process. Soils form by the WEATHERING of material at the soil–parent material interface, or by deposition of sediment on the surface. The parent material is usually rock, but often includes soft sediments. It normally takes 1,000 years to weather 10 cm of material at the B/C horizon interface. By scraping the bedrock and increasing soil aeration and microbial activity, tillage can accelerate the process up to a maximum of about 10 cm in 100 years (i.e. 1 $mm\,yr^{-1}$). If erosion does not exceed this value, the soil system is in a state of dynamic equilibrium. This value has been labelled a 'tolerable' or 'T value' and is used in planning SOIL CONSERVATION strategies.

The 2 $t\,ha^{-1}\,yr^{-1}$ 'T value' may be too high, as it assumes an unlimited supply of weatherable material and ignores the complexity of soil-forming processes. Soils mature through time, usually incorporating organic matter into the topsoil. Erosion preferentially removes topsoil, often with serious implications for soil fertility. The topsoil is the 'seat' of most biological activity and contains most of the soil fauna, organic matter and nutrients, both natural and applied. Therefore, erosion involves more than the loss of physical components of the soil system. Often, the most fertile material is lost and any new soil formed at the base of the profile is much less fertile. The topsoil may be completely stripped away, deposited downslope and then less fertile subsoil or parent material deposited on top. This 'soil profile inversion' diminishes soil fertility. All criticisms suggest the 2 $t\,ha^{-1}\,yr^{-1}$ 'T' value is too high. However, soil erosion rates can exceed it by orders of magnitude and, in extreme cases, completely remove the soil. In the long term, even low erosion rates can be damaging.

Numerous techniques exist for measuring soil losses. For water erosion a simple technique is to establish runoff plots. These are bounded on three sides (usually by wood, metal or plastic) and runoff and eroded sediment are collected at the downslope end. This approach was established in Germany by Ewald Wollny in the 1880s. Runoff and erosion rates are measured in precisely defined conditions of soil type, slope and vegetation cover. However, it has been criticized, mainly because plot boundaries interfere with erosion processes, for example impeding rill development. Runoff plots were nevertheless adopted by the US Soil Conservation Service, following its establishment in 1934. Standardized 0.01 acre plots were constructed in a range of agricultural environments, principally east of the Rockies and led to the development of the 'UNIVERSAL SOIL LOSS EQUATION' (USLE).

Since the USLE was introduced, many soil erosion equations and models have been developed, including the revised USLE (RUSLE) and the Water Erosion Predictive Equation (WEPP). The 'Erosion Productivity Impact Calculator' (EPIC) attempts to predict the long-term effects of erosion on soil properties and crop productivity, simulating erosion for up to hundreds of years. EUROSEM is an attempt to predict erosion rates in European conditions and LISEM is a model to predict erosion on LOESS soils (Plate 128).

Models are also used extensively in wind erosion research. The most notable example is the Bagnold Equation, first published in 1937 by Major R.A. Bagnold. During military service in

Plate 127 The effect of a stone protecting soft silty sediments from splash erosion in the Tabernas Badlands of south-east Spain

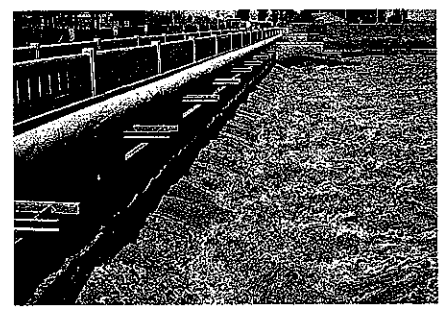


Plate 128 High sediment concentrations in the Yellow River of China. It is estimated that the river transports some 1,800 million tonnes of sediment per year to the Yellow Sea. Much of this sediment is entrained when the river flows through the Loess Plateau, an area of erodible soils derived from wind-blown silt

the British army in Libya, Bagnold studied the dynamics of sand dunes. In the equation, sediment transport is related to wind velocity:

$$Q = 1.5 \times 10^{-9} (V - Vt)^3$$

where Q = total sediment load ($t\,m^{-1}\,h^{-1}$), V = velocity at measuring height ($m\,s^{-1}$), and Vt = fluid threshold velocity for sand movement ($m\,s^{-1}$).

Vt is the wind velocity necessary to initiate effective sand transport. Once that velocity is exceeded, sediment transport increases as the cubic power of wind velocity, so slight increases in wind velocity above Vt cause marked increases in wind erosivity. The Bagnold Equation has formed the basis of other predictive equations, especially in North America. Morgan (1995) comprehensively reviewed erosion models.

There are many techniques to monitor erosion processes in the field and laboratory (De Ploey and Gabriels 1980). A simple technique is to insert erosion pins vertically into the soil, allowing accurate measurements of changing surface levels to show how much erosion (surface lowering) or accumulation (surface raising) has occurred. Simple traps, often referred to as Gerlach 'troughs', can be used in the field to collect sediment. The dynamics of sediment movement can be studied by 'tagging' soil particles with a tracer and using tracer movement to indicate soil movement. Tracers have included painted stones and soil particles, fluorescent dyes, radioactive isotopes and magnetic materials. The fallout of isotopes from nuclear explosions or accidents, such as Chernobyl, has been used to quantify erosion rates. The most widely used isotopes are Cs^{137} and Cs^{134}. These are positively charged and, as they fall to Earth, become attached to negatively charged clays and organic particles. As a site is eroded, it becomes depleted in the isotope so, comparing the radioisotope content of eroded soils with non-eroded soils, such as in a woodland, indicates approximate erosion rates. Another method is to investigate the distribution of fallout isotopes in colluvial deposits. Loughran (1989) summarized field methods to quantify erosion rates. Geomorphologists have also simulated soil erosion processes in the laboratory. However, there are difficulties with this approach, such as how realistically laboratory studies simulate field conditions.

Soil erosion is the product of many complex and interacting factors. For instance, in a laboratory study to assess erosion risk on fifty-five soil samples from the US Cornbelt, Wischmeier and Mannering (1969) found twenty-two soil and surface properties were necessary to explain 95 per cent of soil loss variance. However, Morgan (1995) offered a useful qualitative simplification, stating that soil erosion results from the dynamic interaction of:

1 The energy of the water or wind in causing erosion (EROSIVITY),
2 The inherent resistance of soil to detachment and transport (ERODIBILITY),
3 The protection factor of vegetation.

Many studies have attempted to relate rainfall erosivity to erosion rates. Rainfall erosivity is a function of its intensity and duration and the mass, diameter and velocity of raindrops. The energy of water flowing over the soil surface also affects erosion and is related to its velocity, volume, turbulence and shear stress (Plate 129). Slope angle, length and shape profoundly affect the erosivity of running water. Generally, as slope angle increases, so does erosion risk, because runoff velocity and energy increase. Usually, erodible soils on slopes >10° are particularly susceptible to rill erosion. Rills are hydraulically efficient systems for transporting soil and promote high erosion rates. Increased slope angle also increases the efficacy of splash erosion processes. On longer slopes, runoff has more time to accelerate and thus achieve greater erosivity. Slope shape is also important, as runoff tends to accelerate rapidly over convex slope segments, achieving higher velocities and thus greater erosivity, but decelerates over concave slope sections, often leading to sediment deposition. Thus, the typical convex–concave morphology of slopes in humid temperate environments makes them particularly vulnerable to water erosion.

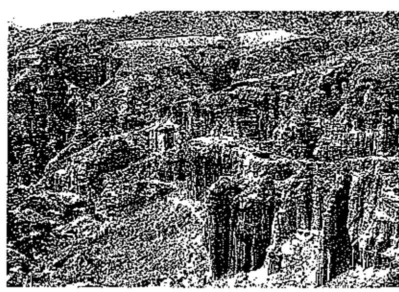


Plate 129 Tu Lin (the 'Soil Forest') in Yunnan Province, China. Highly erosive summer monsoonal rains have incised into soft Tertiary sediments. Human agency has also played a role in promoting erosion. For instance, in the top-centre note the cultivation of melons right at the edge of the gullied area

Soil erodibility is influenced by many factors, principally texture and soil organic content. The most erodible materials are silts and sands; which are not cohesive and can be transported by the flow rates characteristic of rills. Soils with low organic matter contents are highly erodible, particularly those with < 2 per cent organic matter by weight. As organic matter increases above 2 per cent, soil erodibility decreases to a minimum at about 10 per cent organic matter, a typical value for a deciduous forest topsoil in a humid temperate environment, such as the British Isles. Soils with >20 per cent organic matter tend to be more erodible, as there is less clay to form aggregates with organic matter, and organic material is very light and easily transported. Organic particles have densities typically about $0.8\,g\,cm^{-3}$ compared with $>2.5\,g\,cm^{-3}$ for most mineral particles. Low density is the main reason why very organic soils, such as the peaty soils in the Fens of East Anglia, are susceptible to wind erosion.

Vegetation protects the soil from erosive forces by reducing, braking and filtering runoff. Many studies agree that > 30 per cent plant cover protects the soil from erosion. Vegetation surfaces dissipate raindrop energy and prevent slaking. This is particularly true for short vegetation. A fall of about 8–9 m is required to achieve terminal velocity, so raindrops falling from tree canopies can regain their erosivity.

References

Brown, L.R. (1984) The global loss of topsoil, *Journal of Soil and Water Conservation* 39(3), 162–165.

De Ploey, J. and Gabriels, D. (1980) Measuring soil loss and experimental studies, in M.J. Kirkby and R.P.C. Morgan (eds) *Soil Erosion*, 63–108, Chichester: Wiley.

Langbein, W.B. and Schumm, S.A. (1958) Yield of sediment in relation to mean annual precipitation, *Transactions of the American Geophysical Union* 39, 257–266.

Loughran, R.J. (1989) The measurement of soil erosion, *Progress in Physical Geography* 13(2), 216–233.

Morgan, R.P.C. (1995) *Soil Erosion and Conservation*, 2nd edition, London: Longman.

Parrington, J.R., Zoler, W.H. and Aras, N.K. (1983) Asian dust: seasonal transport to the Hawaiian Islands, *Science* 8 April, 195–197.

Wischmeier, W.H. and Mannering, J.V. (1969) Relation of soil properties to its erodibility, *Soil Science Society of America Proceedings* 33, 131–137.

MICHAEL A. FULLEN AND JOHN A. CATT

SOIL GEOMORPHOLOGY

Soil geomorphology is the scientific study of the processes of landscape evolution and the influence that these processes have on soil formation and distribution on the landscape. Soil geomorphology provides a unique framework for an integrated Earth surface assessment. This discipline couples knowledge from soils, surficial deposits, stratigraphy and sedimentation, and parent material with the process-oriented approach of geomorphology in a three-dimensional framework. The soil geomorphology discipline in the United States has provided the primary foundation for interpreting the relation of soils and palaeosols to the landscape.

Soil geomorphology, an interdisciplinary construct, merges two scientific fields: pedology, the study of soils, and geomorphology, the study of landforms and surficial processes. Robert V. Ruhe was one of the first to both quantify landscape form and process in space and time and to integrate these concepts with soil science (e.g. Ruhe 1956, 1960, 1969). Olson (1989, 1997) describes the event chronology leading to the emergence of soil geomorphology as an independent discipline.

Soil geomorphology provides the framework for understanding the geomorphic history of a landscape. A soil-geomorphic approach requires three components: (1) knowledge of the surficial stratigraphy and parent material, (2) geomorphic surfaces defined in time and space and (3) correlation of soil patterns and properties to landscape features. These three components are assessed independently and their results subsequently integrated to produce a soil-geomorphic interpretation or soil-geomorphic model of the landscape and the processes leading to its evolution. A soil-geomorphic model can represent the landscape at defined scales from the hillslope to the continental.

Soil landscape models

A common ground is required to study and understand landscape evolution and the processes that shape the land surface and its soils. The foundation for our current understanding of soils and landscapes lies in the historic concepts of William Morris Davis and Walther Penck. The 'CYCLE OF EROSION' a time-dependent landscape evolution model (Davis 1899), describes landscape progression by downwearing through stages from youth through maturity to old age. In developing this model, Davis was heavily influenced by early theories of evolutionary biology and the concepts of contemporaries including John Wesley Powell and G.K. Gilbert. Slope reduction occurred by uniform downwearing in which geologic differences became insignificant with time as landscapes advanced through the cycle. Process was not a part of this model, a serious flaw. In contrast to the Davisian model, Penck (1924) emphasized backwearing and parallel slope retreat. Penck's model meshes well with our understanding of soil distribution on the landscape and provides a better foundation for soil geomorphology.

Although the Davis and Penck concepts form the basis for general landscape evolution models, they did not emphasize hillslope development in relation to soils. Milne (1936a, b) was one of the first to introduce the CATENA concept to illustrate soil patterns on a hillslope. Here, soil properties depend on topography and are repeated relative to each other from the hillslope summit to the adjacent valley floor. Milne's catena model recognized that the processes of erosion and deposition on the various hillslope positions directly affect the distribution of soil properties. Two variations were recognized: (1) all soils of a catena are formed in a single parent material and (2) soils of a catena are formed in two or more materials. Soils of a catena differ in case (1) because of 'drainage conditions, differential transport and deposition of eroded material and leaching, and translocation and redeposition of mobile chemical constituents' (Milne 1936a). Milne's statement implies that pedogenic processes together with hydrologic properties define the soil landscape and are not restricted to a point on the landscape. In the second case (2), a geologic factor for multiple-parent materials is included. Milne's catena model has evolved into a more limiting model known as a toposequence. Numerous studies have continued to follow Milne's catena model in evaluating soil landscape relations today. The importance of the catena concept to soil science was recognized quickly in the USA. In the USA Soil Survey today, the toposequence is in practice a hydrosequence, i.e. a series of soil profile colours used as indicators of water-table elevation changes along a hillslope.

Wood (1942) and King (1953) created 'the fully developed hillslope' model. Ruhe (1956, 1960, 1975) formulated a soil-geomorphic hillslope model that integrated soil properties with hillslope models and modified the Wood and King models by proposing hillslope elements: summit, shoulder, backslope, footslope and toeslope. These elements are now widely used to describe hillslope positions. Conacher and Dalrymple (1977) extended the two-dimensional hillslope approach to a drainage basin. This nine-unit model, based on form and geomorphic and pedogenic processes, attempted to integrate the components of a hillslope by considering material and water movement.

Landscape morphology

To fully evaluate landscapes and their relations to soils, an understanding of landscape morphology is important. Minimum parameters are gradient, aspect, and vertical and horizontal curvature. Ruhe (1975) described slope curvature using three components: slope gradient, slope length and slope width. A matrix of nine basic forms was used to represent changes in curvature. Huggett (1975) added surface flow lines to the basic slope shapes and Pennock and others (1987) combined hillslope elements and curvature to identify seven hillslope positions.

Soil landscapes and water movement

Water movement is one of the most important mechanisms in landscape evolution and soil development. Water movement is governed by a complex set of interrelated factors both internal (e.g. soil properties) and external (e.g. climate) to the soil-landscape system. Some of the geomorphic models above included water movement, with an emphasis on overland flow or vertical infiltration at a given hillslope position rather than flow through the landscape. This early emphasis reflected the influence of Horton's studies on OVERLAND FLOW in the 1930s. However, in addition to surface runoff and infiltration, throughflow and groundwater recharge are equally important paths for water flow on the landscape. In recent years, the importance of subsurface water movement and transport through the soil landscape has received more attention. Subsurface-flow landscape models including flownet analysis have become important particularly in wetlands studies.

Soils and geomorphic surfaces

Soils and geomorphic surfaces are closely aligned. A geomorphic surface is that part of the landscape specifically defined in space and time with definite geographic boundaries (Ruhe 1956, 1969; Daniels *et al.* 1971). When a soil-geomorphic study is undertaken, the geomorphic surfaces are delineated based on geomorphic principles. Separately, soils in the study area are mapped. The results are compared and the linkages established. Soil boundaries do not necessarily coincide with geomorphic surface delineations and one or more soils may occur on the same geomorphic surface (Ruhe 1975). The critical point to the interpretations is that the pattern of soils and surfaces should be predictably repeated throughout a drainage basin and represents the landscape as it is at a given time. If three different soils occur on a geomorphic surface on one side slope, this soil complex should recur on similarly situated side slopes throughout the watershed. Geomorphic surfaces represent a time sequence but environment and other factors vary independently of one another through time. The latter phenomena affect the types of soils found on any given geomorphic surface. However, the soils will have a common degree of soil development. This relationship provides a valuable tool for understanding the chronological framework of landscape evolution in a soil geomorphic investigation.

Palaeosols and soil geomorphology

Palaeosols are soils formed on landscapes of the past. Ruhe (1975) defined three types: buried, exhumed and relict soils. Buried palaeosols are soils later covered by younger sediment or rock. Exhumed palaeosols are those buried but later re-exposed on the land surface, and relict palaeosols are those formed on a pre-existing landscape but never buried. As researchers began to demonstrate the close interdependence of soils and landscapes (e.g. Ruhe *et al.* 1967; Ruhe 1969), palaeosols also became indicators for understanding past landscapes. Often, buried soils are known as stratigraphic markers and are especially useful as keys to past environments (e.g. Follmer 1982).

Quantifying soil landscape models

Landscape models have shown that landscapes are predictable and have a large non-random variability component. This non-random component is useful in predicting soils on the landscape if the processes that govern landscape development are quantifiably described. Many new and exciting approaches seeking to capture and quantify soil landscape relations encompass geostatistical techniques, digital terrain models, fuzzy logic, neural networks and expert systems and produce visualizations and interactive, virtual immersive modules previously unavailable (e.g. <http://grunwald.ifas.ufl.edu> and <www.soils.wisc.edu/virtual_museum>).

References

Conacher, A.J. and Dalrymple, J.B. (1977) The nine unit landscape model: an approach to pedogeomorphic research, *Geoderma* 18, 1–154.

Daniels, R.B., Gamble, E.E. and Cady, J.G. (1971) The relation between geomorphology and soil morphology and genesis, *Advances in Agronomy* 23, 51–88.

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

Follmer, L.R. (1982) The geomorphology of the Sangamon surface: its spatial and temporal attributes, in C. Thorn (ed.) *Space and Time in Geomorphology*, 117–146, Boston: George Allen and Unwin.

Huggett, R.J. (1975) Soil landscape systems: a model of soil genesis, *Geoderma* 13, 1–22.

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

Milne, G. (1936a) A provisional soil map of East Africa, *Amani Memoirs* No. 28. Eastern African Agricultural Research Station, Tanganyika Territory.

——(1936b) Normal erosion as a factor in soil profile development, *Nature* 138, 548.

Olson, C.G. (1989) Soil geomorphic research and the importance of paleosol stratigraphy to Quaternary investigations, Midwestern USA, *Catena Supplement* 16, 129–142.

——(1997) Systematic soil-geomorphic investigations – contributions of R.V. Ruhe to pedologic interpretation, *Advances in GeoEcology* 29, 415–438.

Penck, W. (1924) *Die Morphologische Analyse (Morphological analysis of landforms)*, trans. by K.C. Boswell and H. Czech (1953), New York: St Martin's Press.

Pennock, D.J., Zebarth, B.J. and E. deJong (1987) Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada, *Geoderma* 40, 297–315.

Ruhe, R.V. (1956) Geomorphic surfaces and the nature of soils, *Soil Science* 82, 441–445.

——(1960) Elements of the soil landscape, *Transactions of the 7th International Congress of Soil Science* 4, 165–170.

——(1969) *Quaternary Landscapes in Iowa*, Ames: Iowa State University Press.

Ruhe, R.V., (1975) *Geomorphology, Geomorphic Processes and Surficial Geology*, Boston: Houghton Mifflin.
Ruhe, R.V., Daniels, R.B. and Cady, J.G. (1967) Landscape evolution and soil formation in southwestern Iowa, *US Dept. of Agriculture Technical Bulletin* 1,349.
Wood, A. (1942) The development of hillside slopes, *Geological Association Proceedings* 53, 128–138.

Further reading

Birkeland, P.W. (1999) *Soils and Geomorphology*, 3rd edition, New York: Oxford University Press.
Carson, M.A. and Kirkby, M.J. (1972) *Hillslope Form and Process*, New York: Cambridge University Press.
Gerrard, A.J. (1981) *Soils and Landforms, An Integration of Geomorphology and Pedology*, Boston: George Allen and Unwin.
Gile, L.H., Hawley, J.W. and Grossman, R.B. (1981) Soils and geomorphology in the Basin and Range area of southern New Mexico – guidebook to the Desert Project, *New Mexico Bur. Mines and Mineral Resources Mem.* 39.
Hall, G.F. and Olson, C.G. (1991) Predicting variability of soils from landscape models, in M.J. Mausbach and L.P. Wilding (eds) *Spatial Variabilities of Soils and Landforms*, 9–24, Special Publication No. 28, Madison: Soil Science Society of America.
Kirkby, M.J. (ed.) (1978) *Hillslope Hydrology*, New York: Wiley.
Richardson, J.L. and Vepraskas, M.J. (eds) (2001) *Wetland Soils. Genesis, Hydrology, Landscapes, and Classification*, New York: Lewis Publishers.
Ruhe, R.V. (1965) Quaternary paleopedology, in H.E. Wright, Jr and D.G. Frey (eds) *The Quaternary of the United States*, 755–764, Princeton: Princeton University Press.
Ruhe, R.V. and Walker, P.H. (1968) Hillslope models and soil formation I. Open systems, *Transactions of the 9th International Congress of Soil Science* 4, 551–560.
Walker, P.H. and Ruhe, R.V. (1968) Hillslope models and soil formation. II. Closed systems, *Transactions of the 9th International Congress of Soil Science* 4, 561–568.
Wright, H.E. Jr, Coffin, B.A. and Aaseng, N.E. (eds) (1992) *The Patterned Peatlands of Minnesota*, Minneapolis: University of Minnesota Press.
Yaalon, D.H. (ed.) (1971) *Paleopedology: Origin, Nature, and Dating of Paleosols*, Jerusalem: Israel University Press.

SEE ALSO: hillslope, form; hillslope, process; Horton's Laws; hydrological geomorphology; palaeosol

CAROLYN G. OLSON

SOLIFLUCTION

Slow downslope movement of soil mass usually associated with FREEZE–THAW CYCLES and FROST HEAVE, first termed to refer to the movement of saturated unfrozen soil in the Falkland Islands (Andersson 1906). MASS MOVEMENTS of soils in periglacial regions involve localized rapid soil failures, which result in active-layer detachment slides or flows, and more widely operating slow pre-failure movements (Ballantyne and Harris 1994: 114). The latter is referred to as solifluction as a collective term. In a broad sense, solifluction consists of frost creep (SOIL CREEP associated with freezing) resulting from nearly vertical settlement of soils heaved normal to the slope and gelifluction representing downslope displacement of ice-rich soils during thawing (Washburn 1979: 201). The two components often operate together, displacing soils downslope generally at rates of 0.5–10 cm a^{-1}. Where the ground is underlain by PERMAFROST, solifluction occurs within the ACTIVE LAYER and is distinguished from permafrost creep that originates from plastic deformation of frozen debris. The long-term operation of solifluction results in low-relief, gentle slopes having a number of lobes and sheets.

Processes

Frost creep is subdivided into needle-ice creep, diurnal frost creep and annual frost creep in terms of the time span of operation and the vertical extent of movement. Annual frost creep is often accompanied by gelifluction. In cold permafrost regions, a large part of annual frost creep and/or gelifluction may occur near the base of the active layer, producing a plug-like velocity profile.

Diurnal freeze–thaw cycles lead to either needle-ice creep or diurnal frost creep (Figure 158). The former occurs when nocturnal cooling to just below 0°C is followed by daytime thawing. Only the uppermost few grains are heaved by ice needles and resettled on thawing by toppling or rotational movement. The resulting downslope movement is superficial but relatively rapid, accelerating approximately with the second power of the slope gradient.

More intensive nocturnal freezing can produce ice lenses at a depth of a few centimetres, which heave the overlying soil layer normal to the slope. On thawing, the heaved soil resettles vertically or, in reality, with some upslope component due to cohesion. Such a freeze–thaw alternation induces diurnal frost creep, the movement of which is proportional to the slope gradient and shallower than 10 cm. The velocity profile of diurnal frost creep is typically concave downslope, in response to the increasing number of freeze–thaw cycles

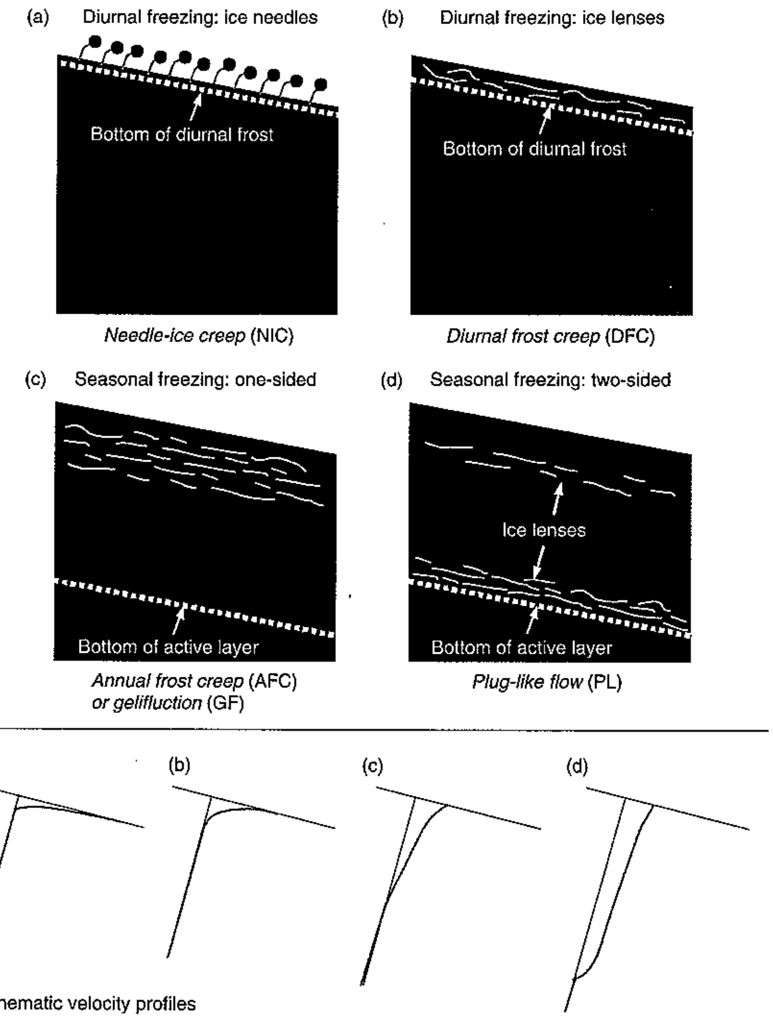


Figure 158 Types of solifluction
Source: Matsuoka (2001)

towards the ground surface. Where needle-ice creep dominates, a gap in velocity occurs between the uppermost grains and the underlying soil (Matsuoka 2001).

Annual freeze–thaw cycles cause frost heave of soils either by one-sided freezing or by two-sided freezing. In regions subjected to seasonal frost or underlain by warm permafrost, one-sided freezing from the ground surface in winter often produces ice lenses within the uppermost few decimetres of soil. During seasonal thawing, the resettlement of the heaved soil results in annual frost creep and/or gelifluction. Reflecting the depth of the ice lenses, the base of movement typically lies at 30–50 cm depth. The relative contribution of gelifluction is estimated by subtracting the potential frost creep, which is the product of the amount of frost heave and the tangent of the slope gradient, from the total downslope displacement. The gelifluction component increases with the silt plus clay content, because excess pore water during thawing reduces frictional strength in the soil (Harris 1996).

Where the permafrost temperature is low, winter freeze-back of the active layer may progress both downward from the top and upward from the base. Such two-sided freezing may produce ice lenses both near the surface and near the permafrost table. The high moisture availability at the latter location favours the formation of numerous thick ice lenses. Thawing of the basal ice lenses induces plug-like flow that shows a velocity profile convex downslope near the base of the active layer (Mackay 1981).

Rates

Field measurements in a wide range of cold regions indicate that rates of solifluction, expressed by either the surface velocity or the volumetric velocity, significantly vary with climate (Matsuoka 2001). In the polar, cold permafrost regions, where the mean annual air temperature (MAAT) is $-6\,°C$ or lower, the paucity of diurnal frost creep restrains the surface velocity to below $5\,\mathrm{cm\,a^{-1}}$, while plug-like flow allows movement of the soil mass deeper than 50 cm. In the alpine, shallow seasonal frost regions, the predominance of diurnal frost creep (including needle-ice creep) raises the surface velocity up to $100\,\mathrm{cm\,a^{-1}}$, whereas the soil movement is confined largely within the uppermost decimetre; as a result, the volumetric velocity is very low despite the high surface velocity. The volumetric velocity reaches a maximum in regions with warm permafrost or deep seasonal frost (MAAT between -6 and $0\,°C$), which are located in both subpolar and alpine settings, because diurnal and annual processes combine to dislocate a 30–50-cm thick soil mass with moderate surface velocities.

The slope gradient is another significant control on solifluction rates. Increasing rates with inclination have been reported from a number of polar slopes. In lower latitude alpine areas, however, other factors like soil frost susceptibility, moisture distribution and freeze–thaw frequency obscure the overall dependence of solifluction rates on inclination (Harris 1981: 123–125). Gelifluction can occur on gradients as low as 1°.

Radiocarbon ages of organic materials buried by solifluction deposits show that long-term variation in solifluction rates generally corresponds to climate change during the Holocene, although the responsible climatic factors are not unequivocal (Matthews *et al.* 1993).

Landforms

Solifluction produces characteristic surface features involving lobes, terraces and sheets. Where vegetation is sparse or absent, sorted stripes (a kind of PATTERNED GROUND) often develop on the tread of these features. The most widespread features are lobes 2–50 m in both width and length. Lobes typically occur in alpine regions where heterogeneous surface conditions localize soil movement, whereas sheets, which lack lateral margins, dominate polar slopes that experience more uniform movement (Plate 130). The downslope edge of these features terminates in a steep riser 0.2–2 m in height, where vegetation, coarse debris or downslope declination acts as a brake on soil movement. As a result, the riser is commonly turf- or stone-banked (Benedict 1970). Terraces may develop under the interaction between wind, vegetation and soil movement (Ballantyne and Harris 1994: 261–267). Isolated boulders on slopes subject to solifluction often move as PLOUGHING BLOCKS AND BOULDERS.

The height of lobes reflects the maximum depth of soil movement. The predominance of shallow diurnal frost creep results in small lobes about 0.2 m in height. Such lobes occur mainly on alpine near-crest locations where thin REGOLITH can only respond to diurnal freeze–thaw action (Matsuoka 2001) or on tropical high mountain slopes where the lack of seasonal variation in temperature highlights the effect of diurnal freeze–thaw action (Bertran *et al.* 1995). Larger and higher lobes develop where solifluction originates mainly from annual freeze–thaw action. The length of lobes depends partly on the time of process operation. Well-developed lobes require several thousand years of activity (Matthews *et al.* 1993).

Plate 130 Turf-banked solifluction lobes on a limestone slope, Swiss National Park (2,400 m ASL)

Numerical simulations suggest that long-term predominance of solifluction, accompanied by subsurface debris production and comminution due mainly to frost weathering, eventually leads to gentle convex-upward slopes, often referred to as CRYOPLANATION surfaces, near mountain crests with possible occurrence of summit TORS (Anderson 2002).

References

Anderson, R. (2002) Modeling the tor-dotted crests, bedrock edges, and parabolic profiles of high alpine surfaces of the Wind River Range, Wyoming, *Geomorphology* 46, 35–58.

Andersson, J.G. (1906) Solifluction, a component of subaerial denudation, *Journal of Geology* 14, 91–112.

Ballantyne, C.K. and Harris, C. (1994) *The Periglaciation of Great Britain*, Cambridge: Cambridge University Press.

Benedict, J.B. (1970) Downslope soil movement in a Colorado alpine region: rates, processes, and climatic significance, *Arctic and Alpine Research* 2, 165–226.

Bertran, P., Francou, B. and Texier, J.P. (1995) Stratified slope deposits: the stone-banked sheets and lobes model, in O. Slaymaker (ed.) *Steepland Geomorphology*, 147–169, Chichester: Wiley.

Harris, C. (1981) *Periglacial Mass-Wasting: A Review of Research*, BGRG Research Monograph 4, Norwich: Geo Abstracts.

——(1996) Physical modelling of periglacial solifluction: review and future strategy, *Permafrost and Periglacial Processes* 7, 349–360.

Mackay, J.R. (1981) Active layer slope movement in a continuous permafrost environment, Garry Island, Northwest Territories, Canada, *Canadian Journal of Earth Sciences* 18, 1,666–1,680.

Matsuoka, N. (2001) Solifluction rates, processes and landforms: a global review, *Earth-Science Reviews* 35, 107–134.

Matthews, J.A., Ballantyne, C.K., Harris, C. and McCarroll, D. (1993) Solifluction and climatic variation in the Holocene: discussion and synthesis, in B. Frenzel (ed.) *Solifluction and Climatic Variation in the Holocene*, 339–361, Stuttgart: Gustav Fisher Verlag.

Washburn, A.L. (1979) *Geocryology: A Survey of Periglacial Processes and Environments*, London: Edward Arnold.

SEE ALSO: freeze–thaw cycle; frost heave; periglacial geomorphology; mass movement

NORIKAZU MATSUOKA

SOLUBILITY

Minerals, and the chemical elements which constitute them, vary in the degree to which they dissolve in water. During the dissolution process, the constituents of the chemical compounds in minerals split up, or dissociate, into water. It follows that the solubility of such compounds can be assessed by measuring the concentrations of the constituent elements in the water after a period of time. Minerals generally display increased solubility with higher temperatures and with increasing acidity, notable exceptions to the latter being silica (as quartz) which increases in solubility above pH 10 and aluminium (as gibbsite or kaolinite) which is least soluble at around pH 6 but increases both as pH rises and falls from this value. The solubility of gypsum (as measured by Ca in solution) does not vary with pH. For gypsum, the rate of water flow is a governing factor.

Maximum concentrations are reached asymptotically over time but nevertheless some practical measures of solubility can be found. The Nernst equation describes the dissolution trend over time:

$$C = C_{max}\,(1 - e^{-kt})$$

where C is concentration at time t, C_{max} = maximum concentration and k is a rate constant for each solute.

Highly soluble compounds have a high value of k and exhibit steep rise in concentration over time. The process of dissolution involves the redistribution of energy – and for dissolution to occur there must be a decrease in free energy – and if a solute and a solvent are composed of similar molecules in terms of structure and electrical properties, then high solubility is favoured (see further Davidson 1978: 83–85).

Water flow rate is also important in determining the solubility of minerals in the field and there are also many other factors which limit the transfer of *in vitro* observations of solubility to the field (Casey *et al.* 1995). For water flow, the rate of dissolution of the more rapidly dissolving minerals becomes transport-limited as the products of dissolution readily accumulate and further dissolution is then only facilitated by the removal of weathering products and the arrival of further chemically unsaturated water. For slowly dissolving minerals, the rate of dissolution tends to be slow as compared to the flow of water and so the overall solutional losses tend to become rate-limited. Thus water flow rates, and, for bare surfaces, duration of rainfall and thus time of wetting, become important controls on solution rates.

Additionally, laboratory measurements of solubility are usually made under controlled conditions, whereas field conditions tend to be variable and possess different combinations of factors which are in practice difficult to separate, like freezing, thawing, salt and biological weathering as well as dissolution (Trudgill and Viles 1998).

References

Casey, W.H., Banfield, J.F., Westrich, H.R. and McLaughlin, L. (1995) What do dissolution experiments tell us about natural weathering? *Chemical Geology* 105, 1–15.

Davidson, D. (1978) *Science for Physical Geographers*, London: Arnold.

Trudgill, S.T. and Viles, H.A. (1998) Field and laboratory approaches to limestone weathering, *Quarterly Journal of Engineering Geology* 31, 333–341.

STEVE TRUDGILL

SOLUTE LOAD AND RATING CURVE

All natural waters contain organic and inorganic material in solution; these are called solutes. Spatial and temporal variation in solute content has been studied as a means of investigating CHEMICAL WEATHERING processes, rates of chemical denudation, evaluating nutrient cycling and elucidating the variable pathways followed by water through the drainage basin. The major solutes of interest are calcium, magnesium, sodium, potassium, chloride, bicarbonates, sulphate, nitrate and silica.

According to the classification of Gibbs (1970), the major natural mechanisms controlling world surface water chemistry are: (a) atmospheric precipitation, both composition and amount; (b) rock weathering; and (c) evaporation and fractional crystallization. When he plotted total dissolved solids (t.d.s.) in milligrams per litre against the weight ratio of sodium to sodium plus calcium for the major rivers of the world, he showed three domains: (1) high t.d.s. and high Na to Na + Ca ratio, where the evaporation/crystallization processes are dominant; (2) average t.d.s. and low Na to Na + Ca ratio; and (3) low t.d.s. and high Na to Na + Ca ratio. From this classification, climate and hydrology are the dominant controls of solute concentration at high and low t.d.s. values and lithology becomes dominant at average t.d.s. values. Meybeck (1988) has provided a more detailed breakdown of river water solutes according to climate, hydrology and relief, on the assumption that the residence time of water in a basin will depend strongly on relief.

Meybeck (1987) estimated the global average chemical denudation rate for crystalline igneous and metamorphic rocks and sandstones and shales at 18–19 tonnes/km^2/year and volcanic rocks at 1.5 times higher. Denudation rate for carbonate rocks is 100 and for evaporites is 423 tonnes/km^2/year. This is in the context of a world average of 42.

The effect of relief on chemical denudation rates varies with geology as carbonate rocks dissolve easily regardless of local topography. No effect of hillslope steepness or extent of recent glaciation has been detected on chemical weathering fluxes in small granitoid watersheds. This implies that physical erosion rates are not critical in influencing chemical weathering of silicates and, by further implication, geology and climate are probably more important than relief on a global basis.

There are further factors that are important. Of these, the anthropogenic and the biotic factors are the most crucial. Vegetation increases chemical weathering by supplying carbon dioxide and organic acids to the soil and it increases water contact time with minerals in the soil by retaining moisture and by locally accelerating water recycling via evapotranspiration-enhanced rainfall. Vegetation also affects weathering by stabilizing soil against erosion and thereby increasing weathering rates in regions of high physical erosion (Berner and Berner 1996). Nutrients, organic carbon and dissolved trace elements will be most affected by anthropogenic and biotic controls, and will, at times, vary quite independently from the topographic, geologic, climatic and hydrologic factors.

The most geomorphically relevant ways of analysing solute data depend on the scale of interest. At individual site and slope scale, diffusive and convective equations have been used to model variations in solute flux (e.g. Carson and Kirkby 1972; Berner 1978), but it is probable that the buried tablet technique (Trudgill 1977) will provide the most reliable relative weathering rate information. At river basin scale, input–output budgets and solute budgets, combined with variable runoff source analysis, are the most popular (e.g. Zeman and Slaymaker 1978; Laudon and Slaymaker 1997). At global scale, Meybeck (1982, 1988) has made major contributions through the use of solute budgets.

Precipitation inputs to the land surface contain solutes in dry and wet fallout. The magnitude and composition of this solute content varies with distance from the ocean. As water moves through the vegetation canopy, the soil and the rock of the drainage basin, the solute concentration and composition will change. Solute concentrations within the soil are influenced by precipitation, interactions with the soil matrix, release of solutes through chemical weathering and biotic uptake and release of nutrients.

The solute content of stream flow will therefore reflect the characteristics of the upstream basin, including its geology, topography, vegetation cover and the variable pathways and residence time associated with water movement through the basin. Concentrations will vary with time in response to hydrologic conditions and will frequently exhibit a dilution effect during storm runoff events.

The precise hydrologic pathway has important implications for residence time and hence on solute enrichment. Identification of the pathway gives an understanding of the magnitude and timing of solute fluxes from different hydrologic reservoirs in the landscape and is therefore essential for the understanding of variations of the stream water chemistry. Hydrograph separation of rain-driven storm flow into its storm and pre-storm components can be carried out by collecting stream water samples for stable isotope analysis before, during and after a storm runoff event. Four assumptions are made: (1) rain isotopic content can be characterized by a single isotopic value; (2) pre-storm component, ground water and vadose water can be characterized by the base flow with a single isotopic value; (3) isotopic content of precipitation will be significantly different from pre-storm runoff values; and (4) contribution of stored surface water to the stream is negligible. Under certain hydrologic conditions, alternative hydrologic tracers (such as silica and electrical conductivity) can be used. The specific advantage of electrical conductivity is that it can be continuously monitored and stored in dataloggers.

Measurements of the solute input into a drainage basin and the output in streamflow provide a means of establishing a solute budget for the basin. The net solute yield (output–input) reflects the production of solutes within the basin. This production may be related to chemical weathering, the uptake of carbon dioxide by weathering reactions and the mineralization of organic material. On a global basis, approximately 50 per cent of the solutes found in river water are the products of chemical weathering, but this value is highly variable depending on lithology.

Rating curves are used to describe relations between solute transport and water discharge. If the rating curve is stable, the water discharge can then be used to predict solute concentration and load. The characteristics of these plots, including slope, degree of scatter and intercept are frequently used to characterize the solute response of a drainage basin.

References

Berner, E.K. and Berner, R.A. (1996) *Global Environments*, Upper Saddle River, NJ: Prentice Hall.

Berner, R.A. (1978) Rate control of mineral dissolution under Earth surface conditions, *American Journal of Science* 278, 1,235–1,252.

Carson, M.A. and Kirkby, M.J. (1972) *Hill Slope Form and Process*, Cambridge: Cambridge University Press.

Gibbs, R.J. (1970) Mechanisms controlling world water chemistry, *Science* 170, 1,088–1,090.

Laudon, H. and Slaymaker, O. (1997) Hydrograph separation using stable isotopes, silica and electrical conductivity: an alpine example, *Journal of Hydrology* 201, 82–101.

Meybeck, M. (1982) Carbon, nitrogen and phosphorus transport by world rivers, *American Journal of Science* 282, 401–450.

——(1987) Global chemical weathering of surficial rocks estimated from river dissolved loads, *American Journal of Science* 287, 401–428.

——(1988) How to establish and use world budgets of riverine materials, in A. Lerman and M. Meybeck (eds) *Physical and Chemical Weathering in Geochemical Cycles*, Dordrecht: Kluwer.

Trudgill, S.T. (1977) Problems in the estimation of short-term variations in limestone erosion processes, *Earth Surface Processes and Landforms* 2, 251–256.

Zeman, L.J. and Slaymaker, O. (1978) Mass balance model for calculation of ionic input loads in atmospheric fallout and discharge from a mountainous basin, *Hydrological Sciences Bulletin* 23, 103–117.

Further reading

Caine, N. (1992) Spatial patterns of geochemical denudation in a Colorado alpine environment, in J.C. Dixon

and A.D. Abrahams (eds) *Periglacial Geomorphology*, 63–88, Chichester: Wiley.
De Boer, D.H. and Campbell, I.A. (1990) Runoff chemistry as an indicator of runoff sources and routing in semi-arid badland drainage basins, *Journal of Hydrology* 121, 379–394.
Paces, T. (1986) Rates of weathering and erosion derived from mass balance in small drainage basins, in S.M. Coleman and D.P. Dethier (eds) *Rates of Chemical Weathering of Rocks and Minerals*, 531–550, Orlando: Academic Press.
Velbel, M.A. (1986) The mathematical basis for determining rate of geochemical and geomorphic processes in small forested watersheds by mass balance, in S.M. Coleman and D.P. Dethier (eds) *Rates of Chemical Weathering of Rocks and Minerals*, 439–451, Orlando: Academic Press.

SEE ALSO: chemical denudation; denudation

OLAV SLAYMAKER

SPALLING

Spalling is the peeling off of platy fragments from the surface of rock. The resulting 'spalls' vary from a few centimetres to several metres in scale but their thickness is usually from 1–5 cm. The under surface of spalled fragments may be very irregular. Spalling can be transitional with EXFOLIATION or SHEETING though the latter usually occurs on a much larger scale. The term 'flaking' is also sometimes used.

Spalling can be attributed to several processes. Fundamentally, differential stresses in the outer layer of rock cause separation. The source of this differential stress may be from the growth of salt or ice crystals. Chemical change may be a source of differential stress as secondary minerals are precipitated. These occupy a greater volume and therefore may exert an outwards force on the rock surface. Expansion and contraction due to thermal change (e.g. forest fires or insolation) may also produce sufficient differential stress (Gray 1965). Change of internal stress equilibrium (e.g. due to erosion) may cause separation of plates from an intact rock mass. Spalling may be accompanied by surface alteration or by CASE HARDENING.

Reference

Gray, W.M. (1965) Surface spalling by thermal stresses in rocks, *Proceedings of the Rock Mechanics Symposium, Toronto*, Department of Mines and Technical Surveys, Ottawa, 85–106.

Further reading

Ollier, C. (1984) *Weathering*, London: Longman.
Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

DAWN T. NICHOLSON

SPELEOTHEM

'Speleothem' (Greek: *spele* – cave, *them* – make) – a general term for minerals precipitated in CAVES. More than 250 different minerals are known (Hill and Forti 1997). Calcite deposited in limestone caves is overwhelmingly predominant, accumulating in both air-filled (vadose) and water-filled (phreatic) conditions. 'Travertine' and 'sinter' are alternative terms, or 'tufa' (see TUFA AND TRAVERTINE) when precipitated on organic frameworks. Aragonite is second in abundance. Gypsum is third, found in gypsum caves and also in limestone caves with gypsum interbeds or where H_2SO_4 can react with the rock. Hydrated carbonates and sulphates (e.g. hydromagnesite, epsomite) are quite common, usually as pastes or powders in small amounts. Other minerals are more localized, associated with particular source conditions in bedrocks or clastic fillings; they include native sulphur, many oxides and hydroxides, halides, nitrates, phosphates, silicates, vanadates and a few organic minerals. There is perennial ice in many cold caves.

$CaCO_3$ (calcite and aragonite) speleothems

Calcite occurs mostly as coarse crystals with c axes oriented to growth ('length fast') or in microcrystalline form with c axes oriented across the direction of growth ('length slow'; see Railsback 2000). Many vadose speleothems display alternations of coarser and finer crystals, often with temporary cessations due to drying or dissolution, zones with dust, mud or organic grains: existing crystals may grow through these or new crystals form upon them. Subaqueous deposits are more homogeneous. Pure calcite is translucent, or opaque white due to fluid inclusions. Colour banding (yellow, brown, red-brown) is common, due chiefly to incorporation of fulvic acid chromophores from soil waters (van Beynen *et al.* 2001). Metals such as iron (red), copper (blue) and nickel (bright green) also provide colour, but rarely in visible concentrations (Cabrol and Mangin 2000).

Aragonite generally occurs as needle-like clusters or massive stalagmitic aggregates. Its deposition instead of calcite is attributed to enrichment of Mg^{2+} ions in the feedwater, due to presence of dolomite ($CaMg{\cdot}2CO_3$) or to evaporative effects. Aragonite speleothems are more common in warmer climates, e.g. wet–dry seasonal alternations of calcite and aragonite are reported in Botswana.

Vadose speleothem shapes are created by gravity, or by growth and capillary forces. Principal gravity types are dripstones (stalactites, stalagmites), and flowstone sheets on floors and/or walls. A 'column' is a stalactite–stalagmite pair grown together.

The fundamental form is the 'straw' stalactite, a monolayer crystal sheath enclosing a feedwater canal and growing downwards only. Leakage from the canal may overplate the sheath, creating tapered (carrot-like) stalactites up to one metre in diameter and several in length. Accelerated deposition on protruberances can add a myriad of subsidiary forms such as crenulations, corbels, drapes and lesser stalactites. 'Curtains' grow downwards where feedwater trickles down a sloping ceiling.

The most simple stalagmite is a 'candlestick' adding all new growth at the top under a nearly constant drip. Varying drip or greater fall height causes terraced or corbelled thickening; at the extreme the form is like a pile of soup plates. More common are conical or tapered forms, broadening into domes with flowstone sheets around them. Some stalagmites are >30 m high and domes may be 50 m or more in diameter.

Flowstones are deposited from film flow and accrete roughly parallel to the host surface. They may extend tens to hundreds of metres downstream of their sources and accumulate to thicknesses of several metres. 'Gours' or 'rimstones' are dams building upwards from irregularities in stream channels or on flowstone surfaces. The greatest impound water to depths of several metres. Rims are often strikingly crenulated.

'Helictites' or 'excentrics' grow where crystal or capillary forces predominate, skewing c-axes to create narrow, curvilinear tubes extending out, up and down from rock or parent stalactites, etc. Most are short, <10–20 cm in length. Dense clustering can form tangled masses like the Medusa's hair. 'Anemolites' grow upwind into prevailing drafts. Clusters of needles fanning outwards ('frostwork') are the principal aragonite excentrics.

'Cave pearls' are spheroidal accretions about a nucleus such as grit agitated by water dripping into a pool. 'Popcorn' ('cave coral') describes semi-spherical accretions on flowstone or other surfaces, often in dense, multilayered clusters.

Subaqueous calcite may precipitate from thermal or meteoric waters. The principal forms are spar linings, e.g. most of the 150+ km of passages in Jewel Cave, South Dakota. Deposition extends from the water table to a limiting depth determined by pressure and Ca^{2+} saturation state. Aggregate thicknesses of one metre or more are known. More complex crystal structures and rounded microcrystalline 'clouds' form in static pools. Water surfaces are marked by shelfstone around the edges and floating rafts of calcite accreted to dust particles.

Distribution and abundance

Speleothems can occur as isolated individuals, in clusters, aligned along fractures, or broadcast. Density can increase until all surfaces are covered. Vadose speleothems grow most readily at shallow depths beneath soils rich in CO_2 in tropical and temperate conditions that permit year-round deposition; the largest individuals and greatest densities are found in these settings.

Many caves have several levels. Speleothems are often fewer and smaller in the lower levels, which are usually younger. In any setting speleothem deposition is largely prohibited where there are impermeable beds, e.g. shales, above a cave.

Growth rates, age and environmental studies

Under optimal conditions (large excess of Ca^{2+} ions, high drip rate and evaporation) straws may extend several centimetres in one year. Normal accretion rates in other speleothems probably range between ~1.0 mm/10^3 yr in cold climate flowstones to >1.0 m/10^3 yr in warm cave entrances. Some speleothems grow at constant rates while others vary by factors of ten or more. Many contain hiatuses caused by drought, cold or change of groundwater routing.

Many speleothems can be dated accurately (+/−1 per cent error) by the $^{230}Th/^{234}U$ method if they are less than ~550 kyr in age. Variations of $^{18}O/^{16}O$ isotope ratios during growth may indicate palaeotemperature changes and $^{13}C/^{12}C$ ratios suggest changes of vegetation amount or type. Where present, annual or event banding revealed by u/v fluorescence and other techniques now permits very high resolution reconstructions of past conditions above caves (Hill and Forti 1997: 271–284).

Gypsum speleothems

Gypsum is deposited in three principal modes: (1) as evaporitic growths within bedrocks or cave sediments, which they rupture – 'evapoturbation'. (2) As scattered encrustations or excentric extrusions on rock, sediments or calcite speleothems. Most frequent are 'flowers', extruded, twisting fibrous bundles up to 50 cm in length. Needles grow from sediments and 'hair' from roofs. Larger, bifurcating stalactites are known in a few caves. (3) As regularly bedded floor or wall encrustations in evaporating pools: thicknesses of several metres occur in Carlsbad Caverns, New Mexico, where much is reprecipitated from alteration crusts formed by H_2SO_4 reacting with the limestone walls.

References

Cabrol, P. and Mangin, A. (2000) *Fleurs de pierre*, Lausanne: Delachaux et Niestle.

Hill, C. and Forti, P. (eds) (1997) *Cave Minerals of the World*, 2nd edition, Huntsville, AL: National Speleological Society of America.

Railsback, L.B. (2000) *An Atlas of Speleothem Microfabrics*, http://www.gly.uga.edu/railsback/speleoatlas/SAindex1.html

Van Beynen, P.E. and Bourbonnière, R., Ford, D. and Schwarcz, H. (2001) Causes of color and fluorescence in speleothems, *Chemical Geology* 175(3–4), 319–341.

DEREK C. FORD

SPHEROIDAL WEATHERING

Spheroidal weathering is the phenomenon whereby joint blocks within the regolith become rounded as a result of the separation of concentric layers of block surfaces. Spheroidal weathering is common in basalt and granite but is also found in dolerites, andesite and some sandstones (Heald *et al.* 1979).

There are two main schools of thought concerning the cause of spheroidal weathering. The first envisages that the separation of shells occurs due to residual stress from cooling and contraction. However, this would not explain the presence of corestones in sedimentary rocks such as sandstones. Ollier (1971) also makes the case that spheroidal weathering is a constant volume process, i.e. there is no accompanying expansion or contraction as is the case with EXFOLIATION and SPALLING.

The second school of thought envisages chemical activity, primarily the process of HYDROLYSIS, leading to migration of mineral elements and their concentration in separate layers. Thus distinct bands of accumulation and depletion become established (Augustithus and Ottemann 1966). Activity is most prevalent at corners and edges and so the tendency is for angular joint-bounded blocks to become rounded, or spheroidal in shape.

Ollier (1984) argues that use of the term spheroidal weathering should be restricted to situations where the weathering front attacks the block from all sides, i.e. the block must be beneath the ground surface. Clearly boulders originally rounded by spheroidal weathering may subsequently become exposed at the surface due to erosion. At this point, the style and rate of weathering are likely to change significantly. If the REGOLITH is subsequently completely removed by erosive agents, landforms known as boulder fields and boulder tors may remain on the surface.

The shape and size of corestones and boulders are determined by the spacing of joints in the intact mass. Spheroidal weathering is likely to be more effective in more closely jointed rocks with wide apertures (gaps between blocks). In this way, the opportunity for permeation of the rock with chemical solutions is optimized.

References

Augustithus, S.S. and Ottemann, J. (1966) On diffusion rings and spheroidal weathering, *Chemical Geology* 1, 201–209.

Heald, M.T., Hollingsworth, T.J. and Smith, R.M. (1979) Alteration of sandstones as revealed by spheroidal weathering, *Journal of Sedimentary Petrology* 49, 901–909.

Ollier, C. (1971) Causes of spheroidal weathering, *Earth-Science Reviews* 7, 127–141.

——(1984) *Weathering*, London: Longman.

Further reading

Chapman, R.W. and Greenfield, M.A. (1949) Spheroidal weathering of igneous rocks, *American Journal of Science* 247, 407–429.

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

Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

SEE ALSO: weathering

DAWN T. NICHOLSON

SPIT

An easily recognized deposition coastal landform, which belies the potential complexity of the formation processes (Gilbert 1890; Davis 1896; Zenkovich 1967; Carter 1988). Spit structure and formation are best analysed through a plan-view perspective, as the coastal configuration in which the feature is developed is crucial to the spit's formation. Spits are found on an irregular coastline where sediment availability and wave power allow a constructional smoothing of the coastline by maintaining open coast longshore beach direction (in the form of a spit) into coastal re-entrants/bays.

Spits are essentially narrow depositional embankment-type features that show a dominance of longshore sediment deposition (growth) over cross-shore sediment movement. A spit's elongation relative to width is an indication of both the coastal sediment availability and net longshore-directed wave-generated transport potential. Such sediment can be from a broad size range, though sand-dominated spits are the most common. Gravel-dominated spits are more likely in mid-upper latitudes where gravel is a major component of coastal sediment availability. As spits are essentially a product of breaking wave activity, mud-dominated spits are unlikely to be observed. A spit's presence generates a back-spit energy lee with low-energy currents (tidal and small wave) and fine sediment stores (tidal banks and marshes).

A spit's plan-view shape is linked to the plan view of breaking-wave crests (and hence longshore transport vectors) determined by near-shore bathymetry. Spits tend to develop where wave refraction cannot accommodate to sudden changes of coastal trend and rapid reduction in breaker approach angle reduces the longshore drift rate to zero at this point. This allows beach deposition to overshoot the directional shift in coastline. The spit builds from this depositional nucleus, its orientation a function of wave refraction accommodating to the changing near-shore bathymetry induced by the presence of the spit.

Spits show a sequence of planform changes that are related to variation in both sediment supply and longshore transport potential, and are best developed when near-shore wave approach is angled along the spit. Spits can occur within a re-entrant when wave crests approach parallel to the re-entrant mouth. The regularity of the spit's plan-view form is a function of wave direction and refraction consistency. A spit is connected to the coast and its proximal sediment source by the neck, while a spit extension occurs at the spit's distal end or terminus. A spit per se, is usually only the subaerial (superstructure) expression of a larger submarine feature (spit platform), the distal position of which is the spit ramp. Ramp deposition controls spit growth and usually has a high fine-sediment proportion, even in gravel-dominated spits, related to wave-generated currents. As most of the sediment for the spit platform is supplied by longshore transport, it mimics sediment availability to the superstructure, though tending towards finer sediment. The spit platform requires an increasing sediment volume as the spit progressively builds into deeper water and as the volume of the superstructure generally remains the same, spit elongation rates will decline over time if the longshore sediment supply rate does not increase. Thus sediment supply rate is a major control on spit development.

There is rapid wave shoaling and landward curvature of the breaking wave crest at the spit terminus given its steep bathymetric gradients. This leads to curvature of the distal structure against the general trend of the spit. Curvature, correlating with decreasing breaker height and sediment fining, is enhanced at times of diminished longshore sediment supply. When supply is reinstated, then the spit can extend in line with its original plan and the recurve is isolated. High volume, but episodic, sediment supply can lead to drift-aligned spits where the spit plan outline is essentially rectilinear despite overlapping recurves (Carter and Orford 1991). This scenario is often associated with the initial formation of spits in a disjointed coastline where sediment supply is formed from isolated finite sediment sources (e.g. a drumlin coastline: Orford *et al.* 1996). Once the spit's sediment source is exhausted then continuing wave power starts to rework the existing spit sediment (cannibalization) leading to a thinning of the spit neck, the increased potential of the superstructure to rollover (as bigger waves break closer to the shore) and the landward movement of the spit into swash-alignment. Cannibalization leads to extension or stabilization of the distal position which then tends to act as a down drift hinge position for control of spit form's swash alignment. Spits generally retreat under a rising relative sea level through overwashing and hence rollover. Retreat evidence is provided by truncated back-spit recurves and back-barrier organics emerging on the seaward face.

Tidal currents can influence the spit terminus, especially when spit growth squeezes coastal inlet width. As inlet tidal hydraulic efficiency increases, any ebb-tidal asymmetry leads to protection of the spit terminus by sand shoals related to delta formation. These in turn influence wave refraction at the terminus and allow swash bars to drive onshore and build up the terminus beach. This can be a source for distal aeolian dunes, even given sediment depletion elsewhere on the spit. Changes to delta deposition due to back-barrier reclamation can affect terminus growth: a decrease in sediment supply can lead to a withering of the spit terminus, thinning of the spit and spit beheading by overwash. The control by ebb deltas on wave refraction patterns can lead to apparent opposing spit growth across tidal inlets despite a dominant single direction of sediment supply. Spits that seal off re-entrants form barriers (see BARRIER AND BARRIER ISLAND).

References

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

Carter, R.W.G. and Orford, J.D. (1991) The sedimentary organization and behaviour of drift-aligned gravel barriers, *Coastal Sediments '91*, American Society of Civil Engineers 1, 934–948.

Davis, W.M. (1896) The outline of Cape Cod, *Proceedings of the American Academy of Arts and Science* 31, 303–332.

Gilbert, G.K. (1890) Lake Bonneville, *US Geological Survey, Monograph* 1, 47–48, Washington.

Orford, J.D., Carter, R.W.G. and Jennings, S.C. (1996) Control domains and morphological phases in gravel-dominated coastal barriers, *Journal of Coastal Research* 12, 589–605.

Zenkovich, V.P. (1967) *Processes of Coastal Development*, 409–447, Edinburgh: Oliver and Boyd.

JULIAN ORFORD

SPRING, SPRINGHEAD

Springs are points where ground water, recharged at higher elevations, emerges at the surface. Depending on the nature of the recharge and of the storage/transmission characteristics of the aquifer through which the water has flowed, they may be permanent (perennial), seasonal or intermittent. In karst areas reversing springs called estavelles are found, particularly in association with poljes. Another common feature of karst areas is the presence of permanent 'underflow' springs and higher, intermittent, 'overflow' springs. Spring flow may show little variation over time or may respond rapidly to recharge, varying over several orders of magnitude. Some springs where the outflow is controlled by a siphoning reservoir system exhibit regular ebbing and flowing with a typical period of minutes to hours. Geysers are periodic hydrothermal springs in which a pressurized body of water is warmed to boiling point and explosive spontaneous boiling occurs as pressure is released.

Springs are found at many elevations from high in mountains to beneath sea level, the vrulja of the Mediterranean being an example of the latter. Spring discharges range over seven orders of magnitude, from seeps to large springs with average flows exceeding $20\,m^3\,s^{-1}$ and instantaneous flows of several hundred $m^3\,s^{-1}$. Most of the largest springs are karstic and only those from fractured volcanic rocks rival their output. The largest is thought to be the Tobio Spring in Papua New Guinea with a mean annual discharge of $85–115\,m^3\,s^{-1}$.

Springs which discharge only percolation water are called exsurgences while those that discharge a mixture of percolation water and water from sinking streams are called resurgences. The term 'rising' is commonly used by speleologists as a synonym for a spring. Those springs where water rises from depth are pressure springs, sometimes termed 'Vauclusian' springs after the Fontaine de Vaucluse in France which has been explored to a depth of 315 m. Where the internal hydraulic head in an aquifer greatly exceeds that required to drive the flow of water, springs exhibit a marked upwelling and are commonly termed 'artesian'.

Karst springs are the output points from a dendritic network of conduits, some of which may be large enough for human exploration (caves). They therefore tend to be both larger and more variable in quantity and quality than springs that emerge from coarse granular or fractured media. The latter may result from the convergence of flow lines in a depression or from the concentration of flow along open fractures such as faults, joints or bedding planes.

Where a spring with a moderate to large discharge has emerged at the same point for a long period a marked springhead or steephead may form. Valleys that begin abruptly at the springhead are termed pocket valleys or reculées. Most are short but some are many tens of metres in height. They may form by headward recession, as water from the spring undermines the rock above it, or by cavern collapse.

Further reading

Ford, D.C. and Williams, P.W. (1989) *Karst Geomorphology and Hydrology*, London: Unwin Hyman.

Jennings, J.N. (1985) *Karst Geomorphology*, Oxford: Blackwell.

LaMoreaux, P.E. and Tanner, J.T. (2001) *Springs and Bottled Waters of the World*, Berlin: Springer-Verlag.

JOHN GUNN

STACK

Stacks are isolated pillars of rock that form when part of a retreating coast is separated from the mainland, usually along joints (see JOINTING) or faults (see FAULT AND FAULT SCARP). Stacks also develop because of folding, tropical KARST submergence, solution pipes, induration and variable rock types. They form in fairly strong rocks with well-defined planes of weakness and are uncommon in weak or thinly bedded rocks with dense joint systems. Stacks can develop from the collapse of arch (see ARCH, NATURAL) roofs, but many form directly from erosion of the cliff face.

Further reading

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

STEP-POOL SYSTEM

Step pools are characteristic bedforms that dominate the channel morphology of steep mountain streams (Chin 1989). Steps are generally composed of cobbles and boulders; they are separated by finer materials forming the pools. Steps and pools alternate to produce a repetitive sequence of bedforms, with a longitudinal profile resembling a staircase. The step-pool morphology similarly develops in bedrock channels and in vegetated basins where channels incorporate woody debris to produce log steps. Step pools are part of a continuum of coarse-grained bedforms that includes POOLS AND RIFFLES (Montgomery and Buffington 1997). Despite external influences, step pools commonly occur with sufficient regularity to produce a rhythmic streambed. They represent a type of meandering in the vertical dimension (Chin 2002).

Step pools are functionally important because they provide hydraulic resistance. Steps induce water to plunge into pools below, promoting tumbling flow where much of the flow's kinetic energy is dissipated by roller eddies. By causing a vertical drop in the water surface elevation as water flows from step to pool, steps also decrease potential energy that otherwise would be available for conversion to a longitudinal component of kinetic energy used for erosion and sediment transport. Steps provide the ability to counteract steep slopes, thereby preventing excessive erosion and channel degradation. The role of step pools is especially important in confined mountain streams where lateral adjustments and energy dissipation by meandering and braiding are prohibited.

Step pools form an integral part of the hydraulic geometry of mountain streams. Consistent relations exist between step length, step height and channel gradient. Step length increases and height decreases with a decrease in slope. Such relations are found regardless of substrate type and the presence of woody debris, suggesting that step characteristics are controlled, at least in part, by flow energy expenditure (Wohl *et al.* 1997). Step pools represent a means of adjusting boundary roughness. They evolve toward a condition of maximum resistance, which is apparently achieved when the ratio of mean step height to mean step length to channel slope is between 1 and 2 (Abrahams *et al.* 1995).

Step pools are mobilized by high-magnitude, low frequency floods on the order of fifty years or more (Grant *et al.* 1990). The specific generating mechanism is incompletely understood. Laboratory experiments suggest that steps may originate as antidunes under high flows (Whittaker and Jaeggi 1982). However, although limited field data support the antidune model, the theory cannot explain step-pool formation in all cases. For example, step pools develop in some channels where flows are unlikely to completely submerge clasts and form antidunes. Alternative explanations focus on flow instabilities and the random movement of large particles.

The step-pool bed configuration controls hydraulics and sediment transport in distinct ways.

For example, velocity and flow resistance fluctuate between step and pool and along with increasing discharge (Lee and Ferguson 2002). Sediment transport is episodic, characterized by alternating transport steps and intervals of non-movement (Schmidt and Ergenzinger 1992). Thus, prediction of sediment transport in step-pool streams using standard equations is problematic. New data from instrumented watersheds (e.g. Lenzi 2001) have the potential to yield considerable insights for bed-load transport and associated step-pool processes.

References

Abrahams, A.D., Li, G. and Atkinson, J.F. (1995) Step-pool streams: adjustment to maximum flow resistance, *Water Resources Research* 31, 2,593–2,602.
Chin, A. (1989) Step pools in stream channels, *Progress in Physical Geography* 13, 391–407.
——(2002) The periodic nature of step-pool mountain streams, *American Journal of Science* 302, 144–167.
Grant, G.E., Swanson, F.J. and Wolman, M.G. (1990) Pattern and origin of stepped-bed morphology in high-gradient streams, western Cascades, Oregon, *Geological Society of America Bulletin* 102, 340–352.
Lee, A.J. and Ferguson, R.I. (2002) Velocity and flow resistance in step-pool streams, *Geomorphology* 46, 59–71.
Lenzi, M.A. (2001) Step-pool evolution in the Rio Cordon, northeastern Italy, *Earth Surface Processes and Landforms* 26, 991–1,008.
Montgomery, D.R. and Buffington, J.M. (1997) Channel-reach morphology in mountain drainage basins, *Geological Society of America Bulletin* 109, 596–611.
Schmidt, K.H. and Ergenzinger, P. (1992) Bedload entrainment, travel lengths, steplengths, rest periods – studied with passive (iron, magnetic) and active (radio) tracer techniques, *Earth Surface Processes and Landforms* 17, 147–165.
Whittaker, J.G. and Jaeggi, M.N.R. (1982) Origin of step-pool systems in mountain streams, *Journal of the Hydraulic Division*, ASCE 108, 758–773.
Wohl, E., Madsen, S. and MacDonald, L. (1997) Characteristic of log and clast bed-steps in step-pool streams of northwestern Montana, USA, *Geomorphology* 20, 1–10.

Further reading

Wohl, E. (2000) *Mountain Rivers*, Washington, DC: American Geophysical Union.

SEE ALSO: gravel-bed river

ANNE CHIN

STERIC EFFECT

An effect in which the molecular dimensions of the material controls the rate or path of a physical or chemical reaction. A steric effect on a rate process may lead to a rate increase (steric acceleration) or a decrease (steric retardation). The most significant steric effect in geomorphology is SEA LEVEL change. In sea water, steric effects are driven by changes in temperature and differences in salinity (i.e. density). As heat is able to exchange freely with the atmosphere, temperature is the dominant steric parameter, particularly over longer timescales (thousands of years) where salinity remains fairly constant in oceans. The change in seawater level as a result of steric effects is referred to as the steric height. This is defined as the height of the sea as the integral of the specific volume from a specified pressure level to the ocean surface. However, steric heights are hard to calculate and trends difficult to attain, as changes occur over varied spatial and temporal scales.

Steric changes occur on seasonal and intra-annual time periods, mostly as a result of steric effects in the upper 500 m of the oceans where heat exchange is much more rapid. It is estimated that an increase in temperature of 1 °C throughout the uppermost 500 m will result in a sea-level rise of approximately 100 mm. In comparison, colder deeper waters show slower heat exchange, and it has been argued that they should be ignored in steric height calculations. However, a parcel of water at 4 °C at 2,000 m depth has a thermal expansion coefficient (which increases with temperature and pressure) 60 per cent as large as that of a parcel at the ocean surface at 20 °C (Roemmich 1990). A warming of the entire sea-water column of 1 °C would raise the sea level by about 0.2 m, whereas a warming of 10 °C would lead to an increase in sea level of about 8 m (as the coefficient of thermal expansion increases with temperature) (Knutti and Stocker 2000).

Steric height variations have been shown to influence global sea-level fluctuations on short-term timescales in particular (seasonal, intra-annual and decadal periods on the order of 10 cm), yet steric effects are not great enough to account for sea-level variations over greater timescales (for instance, the 150 m rise in sea level since the last glacial period) (Roemmich 1990). However, thermal contraction of sea water could account for a sea-level rise between the climatically warm Cretaceous period (144–66 Ma) and the onset of the ice-dominated system in the Cenozoic.

Steric effects hold great contemporary importance with regards to global warming, and several modelling studies have investigated past changes and the likely response of the oceans to future changes (e.g. Knutti and Stocker 2000). A rise in sea level of between 10–50 cm is projected due to steric changes over the next century.

References

Knutti, R. and Stocker, T.F. (2000) Influence of the thermohaline circulation on projected sea level rise, *Journal of Climate* 13, 1,997–2,001.
Roemmich, D. (1990) Sea level and the thermal variability of the ocean, in National Research Council *Sea Level Change (Studies in Geophysics)*, Washington, DC: National Academy Press.

STEVE WARD

STONE-LINE

Stone-lines are synonymous with 'stone-layers', *nappes de gravats* (French) and *Steinlagen* (German) and are common stratigraphic features within many tropical soils and weathering profiles. They form striking, mainly undulating and approximately downslope-oriented discontinuities in soils, consisting of stringers of resistant, largely unweathered coarse clasts at different depths below the groundsurface. In thickness they may range from several centimetres up to a metre or even more. These layers of coarse material separate the overlying loamy to sandy topsoil, or hillwash, from the strongly chemically weathered subsoil or SAPROLITE. In most cases the substrate is not well sorted and may be angular to subangular or possibly well rounded in shape, comparable to gravels and pebbles. Although usually dominated by quartz, pisoliths, iron nodules and larger fragments of lateritic crusts, as well as human artefacts can occur.

A three-stepped stratigraphic subdivision of the stone-line complex was developed for Central Africa, distinguishing the cover (hillwash) or α-layer, the stone-line or β-layer, and the weathered rock (saprolite) or γ-layer (Stoops 1967). The α-layer is typically a few centimetres to some metres thick. It consists of loose but structured material, of sandy to clayey texture, practically devoid of elements coarser than 4 mm (with the exception of some loose iron concretions, mainly at the surface). It shows no stratification. The hillwash follows the general slope of the hill, albeit with a locally more sinuous path with introversions. The sinuosity amplitude varies between 2 and 4 m, and the height-differences seldom reach 50 cm. At the transition from the hillwash (α-layer) to the stone-line (β-layer), slightly coarser material up to 1 cm in diameter occurs. In striking contrast to the covering hillwash, the stone-line often shows a vertical zonation dominated by weathered vein-quartz and rolled pebbles with iron coatings, which are reminiscent of alluvial gravel. Prehistoric implements have been identified in some locations (β_1-layer). Along some sections only, a gradual transition to the lower, so-called β_2-layer takes place. This part of the stone-line consists essentially of fragments of *in situ* weathered quartz veins and chert bands showing a sub-angular to angular shape. Surface coatings of clasts are rare. Thickness of this layer can reach several centimetres to two metres. Below the stone-line a profound weathering zone characterized by kaolinitic clay with a subsequent transition to bedrock (mottled and pallid zone) is recognizable. A further subdivison of the γ-layer relating to soil colour and micromorphological properties, in connection with successive alteration of the soil profile (e.g. γ_1-, γ_2-, γ_3-layer) may be present.

Due to a huge variety of stone-line phenomena in the tropics an extensive literature on these stratigraphic features exists, indicating that several theories and geomorphic processes may be considered in explaining their morphogenetic origin. Early in stone-line research in the late 1940s it was thought that stones, originally dispersed over the whole depth of the soil profile, were able to sink through the matrix of fines, and would finally concentrate on top of the underlying bedrock. However, this conception proved wrong as the bearing capacity of a soil generally remains high enough to support stones. Today, it is generally accepted that stone-line formation is closely linked to a catenary context, especially to the domain of hillslope processes and slope morphology. It is equally related to long-term weathering and morphodynamics of landscapes. Nevertheless the interpretation of different stone-line features, in particular whether they are the result of an *in situ*, autochthonous formation by down-weathering, or whether their origin lies in a combination of laterally active morphodynamic processes due to former palaeoenvironmental modifications of the landscape (allochthonous formation) is still controversial.

Stone-line formation

The following explanations of stone-line formation are most likely:

1. Stone-lines are residual surface accumulations (palaeopavements) which were later covered by finer sediments. This process results from selective erosion by episodic sheet wash (see SHEET EROSION, SHEET FLOW, SHEET WASH), soil creep and the formation of slope pediments with retreating scarps, causing the hillslope-oriented accumulation of colluvial- or hillwash-like fine material. This landscape instability is effected by a climate change to drier, arid conditions (Rohdenburg 1969).
2. Stone-lines occuring near river plains can be understood as parts of palaeochannels (e.g. former anastomosing branches), that were formed by redistribution and concentration of gravel by surface water flows and related colluvial activity.
3. Stone-lines may be considered as the result of bioturbation by termites, ants and worms. Selective zoogenic uptake of fine material from bottom to top in a soil leads to a concentration of coarser material in greater depth.
4. It is most likely that the formation of many stone-lines has to be considered within the context of drier climatic conditions during the Last Glacial Maximum (LGM). However, there is also evidence that stone-lines can be interpreted as stratigraphic markers of a Younger Dryas event with cold and dry climatic conditions at the onset of the Holocene (Runge 2001).

Economic significance of stone-lines

The coarse material is frequently quarried for road construction as paving gravel and for mineral exploration (e.g. cassiterite, columbite, gold, monazite, zircon, rutile, ilmenite, diamonds). Due to their greater specific weight and greater mechanical and chemical weathering resistance these minerals are concentrated in stone-lines (placer deposits). Such sites are often the first to be exploited as no heavy equipment is required. The mineral content of stone-lines is often used as a pathfinder towards major hardrock orebodies (Thorp 1987).

References

Rohdenburg, H. (1969) Hangpedimentation und Klimawechsel als wichtigste Faktoren der Flächen und Stufenbildung in den wechselfeuchten Tropen, *Giess. Geogr. Schriften* 20, 57–152.

Runge, J. (2001) On the age of stone-lines and hillwash sediments in the eastern Congo basin – palaeoenvironmental implications, *Palaeoecology of Africa and the Surrounding Islands* 27, 19–36.

Stoops, G. (1967) Le profil d'altération aus Bas-Congo (Kinshasa). Sa description et sa genèse, *Pédologie* 17, 60–105.

Thorp, M.B. (1987) The economic significance of stone-lines, *Geo-Eco-Trop*, 11, 225–227.

Further reading

Alexandre, J. and Malaisse, F. (eds) (1987) Journée d'étude sur 'Stone-lines', Bruxelles, *Géo-Eco-Trop* 11, 1–239.

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

SEE ALSO: saprolite; sheet erosion, sheet flow, sheet wash

JÜRGEN RUNGE

STONE PAVEMENT

Sometimes called 'desert pavement', a stone pavement is an armoured surface composed of a thin mosaic of rock fragments that is set in or on a matrix of finer material. In Australia the rock fragments may consist of SILCRETE fragments, locally termed *gibbers*. Pavements are important because they are a major control on surface stability. They also provide a record of the activities working on desert surfaces. In addition, if they are disrupted, accelerated erosion of the underlying finer material may occur. Pavements also act as a store for material in transit by the wind and may fundamentally affect infiltration, runoff and sediment erosion rates (Poesen *et al.* 1994). They tend to occur in areas with limited vegetation cover, and the presence of vegetation, as in wet years, can encourage increased levels of bioturbation and surface disturbance (Haff 2001). Stone pavements are not restricted to deserts, however, and occur, *inter alia*, in tundra regions.

Stone pavements may display soil horizons or they may be produced without appreciable soil development, especially in the case of immature examples developed on relatively unstable ground surfaces by superficial processes, such as deflation and runoff. In those examples with soil horizonation, there is often a vesicular A horizon of mainly silt-clay-size particles (McFadden *et al.* 1998).

There has been a great deal of discussion about the processes that lead to pavement formation.

A classic model is that of deflational sorting, whereby a lag of coarse material is left at the surface after finer materials have been removed by wind action. Another possible mode of horizontal removal of fines needs to be considered, however, namely water sorting (Cooke 1970). Upward migration processes may also play a role, for the concentration of coarse particles at the surface and at depth, and the relative scarcity of coarse particles in the upper part of the underlying soil profile, suggests that the coarse particles may have migrated upwards through the soil to the surface. This could be achieved by a range of processes, including freezing and thawing, wetting and drying, changes in salt phases and bioturbation. In addition, pavement characteristics may be much modified by the addition of aeolian materials, especially dust, from above (McFadden *et al.* 1987; Wells *et al.* 1985). Pavement characteristics also change with age, with pavement development being greater on older surfaces (Amit and Gerson 1986). In particular, the nature of older surfaces will be characterized by a greater degree of particle weathering caused by processes like salt weathering. They may also display greater degrees of rock varnish cover.

In recent years there has been great interest in the way that pavements recover from disturbance brought about, for example, by the passage of wheeled transport. In the Western Desert of Egypt, vehicle tracks from the two world wars of the twentieth century can still be clearly seen on pavement surfaces. However, in other localities pavements have been seen to heal relatively rapidly following deliberate local disruption of their surfaces. Haff and Werner (1996), working in California, found that gaps healed in around 5 years and that displacement of surface stones by small animals was a major component of the healing process. Similarly, Wainwright *et al.* (1999), using rainfall simulation experiments at Walnut Gulch in Arizona, found that raindrop erosion processes resulted in rapid surface recovery.

References

Amit, R. and Gerson, R. (1986) The evolution of Holocene (reg) gravelly soils in deserts. An example from the Dead Sea region, *Catena* 13, 59–79.

Cooke, R.U. (1970) Stone pavements in deserts, *Annals of the Association of American Geographers* 60, 560–577.

Haff, P.K. (2001) Desert pavement: an environmental canary? *Journal of Geology* 110, 661–668.

Haff, P.K. and Werner, B.T. (1996) Dynamical processes on desert pavements and the healing of surficial disturbances, *Quaternary Research* 15, 38–46

McFadden, L.D., McDonald, E.V., Wells S.G., Anderson, K., Quade, J. and Forman, S.C. (1998) The vascular layer and carbonate collars of desert soils and pavements: formation, age, and relation to climate, *Geomorphology* 24, 101–145.

McFadden, L.D., Wells, S.G. and Jercinovich, M.J. (1987) Influence of aeolian and pedogenic processes on the origin and activities and evolution of desert pavements, *Geology* 15, 504–508.

Poesen, J.W., Torri, D. Burt, K. (1994) Effect of rich fragments on soil erosion by water at different spatial scales: a review, *Catena* 23, 141–166.

Wainwright, J., Parsons, A.J. and Abrahams, A.D. (1999) Field and complete experiments on the formation of desert pavements, *Earth Surface Processes and Landforms* 24, 1,025–1,037.

Wells, S.G., Dohrenwend, J.C., McFadden, L.D., Turrin, B.D. and Mehrer, K.D. (1985) Late cenozoic landscape evolution on lava flow surfaces of the Cima volcanic field, Mojave Desert, California, *Geological Society of America Bulletin* 96, 1,518–1,529.

A.S. GOUDIE

STORM SURGE

Storm surge is a response of the ocean to changing atmospheric pressure and strong winds caused by cyclonic weather systems, that can result in higher water surface elevations than are predicted by normal astronomical tides, lasting between an hour and four days but typically in the order of 6–18 hours. Storm surge results from the combined action of extreme wind shear stress on the ocean which moves and holds water against windward coasts (wind set-up), and the inverse-barometer effect of changing atmospheric pressure that increases the mean water surface level as pressure drops (pressure set-up). Pressure set-up increases the average water surface by 1 cm per 1 hPa drop in atmospheric pressure, but in cases of extreme storm surge, wind set-up is much more significant. Wind speeds, the track and the relative position of the storm centre to the shore, the slope of the continental shelf and the configuration of the shoreline (particularly the extent of embayment) are all influential in determining the size of the storm surge. The largest storm surges result from hurricanes (otherwise known as tropical cyclones or typhoons) and raised water levels of up to 8 m have been reported. Significant storms at higher latitudes tend to produce surges in the order of 1–3 m, although on shallow continental shelves higher surges are possible.

In generally low-lying coastal areas, increases in mean water level by storm surge can result in large areas of land inundation, often coinciding with floods and other storm-related effects. The most significant regularly affected locations for storm surge damage are the Bay of Bengal, the south-east coast of the USA and the east coast of China. Surges in the shallow, funnel shaped Bay of Bengal have resulted in more than 100,000 deaths on four occasions since 1897 (Bao and Healy 2002), with the worst event occurring in 1970 resulting in the loss of approximately 300,000 lives. In the USA, although surges result in inundation of significant areas of land and high economic loss, deaths have been few due to good warning systems, and disaster response infrastructure. Considerable effort has gone into understanding and modelling storm surge (Bode and Hardy 1997).

The effects of storm surge can be exacerbated if the surge coincides with periods of high astronomical tide. Additionally, on open coasts storm surge is normally associated with increased wave set-up and the establishment of long-period wave motions in the surf zone, all of which increase water level on the beach, allowing storm waves to penetrate much further inland which can lead to significant coastal erosion.

References

Bao, C. and Healy, T. (2002) Typhoon storm surge and some effects on muddy coasts, in T. Healy, Y. Wang and J.-A. Healy (eds) *Muddy Coasts of the World: Processes, Deposits and Function*, 263–278, Amsterdam: Elsevier Science.

Bode, L. and Hardy, T.A. (1997) Progress and recent developments in storm surge modeling, *Journal of Hydraulic Engineering* 123(4), 315–331.

SEE ALSO: continental shelf; overwashing; wave

KEVIN PARNELL

STRANDFLAT

The word 'strandflat' is a name used for the low country and shallow sea along the western Norwegian coast, and also along coasts in Arctic and Antarctic areas that have been covered by ice sheets during the Quaternary ice age. Apart from long stretches of the west coast of Norway where the strandflat is an almost continuous feature, the strandflat has also been recognized in areas as far apart as the South Shetland Isles, Alaska and western Scotland. The low areas of strandflat often appear as broad glacially moulded coastal rock platforms sometimes as much as 80 km in width and backed by high cliffs. However, these shore platforms generally exhibit considerable local relief thus making it difficult to assign a precise altitude to any individual area of platform.

The strandflat was first described by Reusch (1894) while its possible origins were first considered in detail by Nansen (1922). The various processes of strandflat formation are well summarized by Larsen and Holtedahl (1985) and include marine abrasion, subaerial weathering, glacial erosion, frost shattering and cold climate shore erosion. Larsen and Holtedahl proposed that the strandflat was primarily the result of sea-ice erosion and frost shattering during the Quaternary and that most surfaces had been later modified by marine erosional and glacial erosional processes. They also noted that the Norwegian strandflat surfaces exhibit glacio-isostatic tilting and are therefore likely to have been produced during periods of Quaternary glaciation rather than during temperate interglacial periods.

It is difficult to determine precise ages for the formation of the various strandflat surfaces around the world although recent developments in cosmogenic isotope dating techniques represent one possible way forward of dating individual rock surfaces. Consideration of the Quaternary marine oxygen isotope record and of the glacio-isostatic changes that have affected land areas buried by Quaternary ice sheets implies that the position of relative sea level in any area is unlikely to have remained stationary for any significant length of time, certainly no less than *c.*10,000 years. Accordingly, this would seem to point to the conclusion that individual strandflat surfaces having been produced by cold climate shore processes, must have been repeatedly overwhelmed by ice sheets and subject to marine processes during numerous intervals of cold climate throughout the Quaternary.

References

Larsen, E. and Holtedahl, H. (1985) The Norwegian strandflat: a reconsideration of its age and origin, *Norsk Geologiske Tidsskrift* 65, 247–254.

Nansen, F. (1922) The strandflat and isostasy, *Videnskapelkapets Skrifter 1. Math.-Naturw. Kl. (Kristiana)*, 11.

Reusch, H. (1894) Strandfladen et nyt traek I Norges geografi, *Norges Geologiske Underssokelse* 14, 1–14.

ALASTAIR G. DAWSON

STREAM ORDERING

Stream ordering is a technique for characterizing the constituent parts of a drainage network. Ordering can start from the outlet and move in the upstream direction or it can start from each source and move downstream. The most successful have been those ordering systems which move in a downstream direction. The upstream moving systems require a series of subjective decisions about which upstream extension is the master stream. Horton (1932, 1945) introduced the following ordering system:

(a) Channels that originate at a source, and have no tributaries are defined to be first-order streams;
(b) When two streams of order x join, a stream of order $x + 1$ is created;
(c) When two streams of different order join, the channel segment immediately downstream of the junction takes the higher order of the two combining streams;
(d) When the highest order stream segment (n) has been defined, then the upstream extensions of that segment are deemed to have the same order (n) all the way to the source. Similarly, stream segments of order $(n-1)$ are extended back to their source and so on.

This hybrid system (first downstream and then upstream) incorporates the subjectivity mentioned and therefore Strahler (1952, 1957) revised Horton's scheme by removing step (d) above. This so-called Strahler system (or Horton-Strahler ordering system) is the most commonly used in hydro-geomorphology.

A third-ordering system is the link magnitude system proposed by Shreve (1966). Source streams or links have magnitude 1. At a bifurcation, the downstream link takes the magnitude of the sum of the magnitudes of the two incoming links. The magnitude of each link is therefore equivalent to the number of sources in the network draining into that link.

The theoretical basis for Shreve's ordering system is his view of the river basin as a random topological structure. Terminology used includes: a node is the one outlet furthest downstream; n sources are the points furthest upstream and there are $n-1$ junctions. Edges of the network are links; exterior links emanate from sources; interior links emanate from junctions. A network with n sources has $2n-1$ links, n of which are exterior and $n-1$ are interior links.

The most important measuring device that Horton identified was that involving stream ordering. The idea originated with a German hydraulic engineer Gravelius (1914). Chorley (1995) noted that there were two important corollaries of Horton's stream ordering: (a) it placed the emphasis on analysis based on the identification of individual drainage basins. The latter thus emerged as rational, clearly defined topographic units whose geomorphic status was expressed by their order and which prompted geometrical comparisons from one location to another; and (b) the procedure generated a nested hierarchy of drainage basin forms, each of which could be viewed as an open physical system in terms of inputs of precipitation and outputs of discharge and sediment load. A third corollary could be added to Chorley's list. The application of increasingly refined statistical analysis to problems of watershed geomorphology was facilitated by ordering the channels: issues of sample size and representativeness became central considerations in geomorphology.

Perhaps most importantly, the concept of drainage density, which was to become the most important geometric indicator in the work of the Columbia school of quantitative geomorphology, was encouraged by the ordering of streams. Melton (1957) explored the relation between drainage density and stream frequency, as well as the ratio of the two as a measure of the completeness with which a channel system fills a basin outline and as a possible evolutionary index of drainage basins.

Horton's ordering also generated the laws of drainage composition. These were exponential relations between stream order and (a) number of streams of a given order; (b) average length of streams of each order; (c) total stream lengths of each order; (d) basin areas of each order; and (e) average stream slopes of each order. Each of Horton's laws generated a ratio (e.g. the bifurcation ratio from the first law) and these ratios were shown to lie within quite narrow ranges except where differential geological controls were important. Horton's laws provided a topologically and geometrically logical set of procedures for the analysis of fluvially dissected terrain. Much geomorphic work has subsequently been based on systems of stream ordering.

References

Chorley, R.J. (1995) Classics in physical geography revisited: Horton, R.E., 1945, *Progress in Physical Geography* 19, 533–554.

Gravelius, H. (1914) *Flusskunde*, Berlin: Goschensche Verlagshandlung.
Horton, R.E. (1932) Drainage basin characteristics, *American Geophysical Union Transactions* 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.
Melton, M.A. (1957) An analysis of the relations among elements of climate, surface properties and geomorphology, *Office of Naval Research Project NR 389-042, Technical Report 11*, New York: Columbia University Press.
Shreve, R.L. (1966) Statistical law of stream numbers, *Journal of Geology* 74, 17–37.
Strahler, A.N. (1952) Hypsometric (area-altitude) analysis of erosional topography, *Geological Society of America Bulletin* 63, 1,117–1,142.
——(1957) Quantitative analysis of watershed geomorphology, *American Geophysical Union Transactions* 38, 912–920.

SEE ALSO: allometry; drainage density; dynamic geomorphology

OLAV SLAYMAKER

STREAM POWER

Power is the rate of doing work (force × distance) and is expressed in Watts which are Joules per second ($J s^{-1}$). Stream power is the rate at which a stream can do work, especially in the transport of its sediment load, and is usually measured over a specific length of channel. It expresses the rate of energy expenditure in flowing water and, as such, could provide a basic integrating theme within the physical environment (Gregory 1987). In hydraulics and fluvial processes, attempts to analyse the processes involved have used a number of variables whereas expressing energy expenditure as stream power is a more fundamental approach. The potential energy that water possesses at a particular location is proportional to its height above some datum which can be sea level or a lake level; this potential energy is converted into kinetic energy as the water flows downhill under the influence of gravity.

Three important aspects of stream power are: how it is expressed, what controls it, and how has it been utilized and applied. Stream power (ω) was first expressed (Bagnold 1960) as the product of fluid density (ρ), discharge (Q), acceleration due to gravity (g) and slope (s) in the form:

$$\omega = \rho Qgs$$

This expression for power can of course be applied to any fluid, and Bagnold used a similar approach in relation to wind movement over the Earth's surface. Bagnold's (1960) definition has subsequently been applied to the rate of sediment transport (Bagnold 1977) expressed as the amount of energy expended per unit area of the bed. Such unit stream power could be obtained per unit channel width (w) or bed area as:

$$\omega = \frac{\rho Qgs}{w}$$

which, because $Q = wdv$, is simplified to $\omega = \rho gdvs$ thus including depth (d) and velocity (v), and this is often referred to as specific power. Stream power is measured in Joules per second (Js^{-1}) (Watts) and unit stream power is expressed per square metre ($J m^{-2} s^{-1}$ or Wm^{-2}). Unit stream power is expressed per unit length of channel or per unit channel area, and when comparing results from several areas it is important to differentiate between the results achieved by several methods. Values of unit stream power ω range from less than $1\,Wm^{-2}$ in flow between rills to $>12{,}000\,Wm^{-2}$ in riverine flood flows in India, up to $18{,}582\,Wm^{-2}$ for large flash floods, and up to $3 \times 10^5\,Wm^{-2}$ for the Missoula flood (Baker and Costa 1987) in the Quaternary, the largest known discharge of water on Earth.

Controls upon stream power can be deduced from its component elements, of which g is constant and ρ, Q and s are variables. Along a single channel, slope may tend to decrease downstream, whereas discharge will increase, and there can be significant variations in water quality and sediment transport which affect the value of ρ. Along a single river unit, stream power tends to peak in the middle parts of some basins; it is lower upstream, where discharges are relatively small, and lower downstream, where river gradients have their lowest values. Stream power can be calculated for the discharge in the channel at any one specific time or it can be calculated for the estimated channel capacity flow or for some flood flow value; the pattern of stream power distribution down valley may vary somewhat in each case.

Applications of stream power have now been made to many aspects of analysis of river systems. Most usefully in relation to *sediment transport*, stream power has been used instead of stream discharge, velocity or bed shear stress to relate to sediment motion and transport, especially that of bedload (e.g. Allen 1977). This approach is arguably more geomorphological than using hydrological parameters. It has also been used more generally as a means of considering the efficiency of sediment transport; by comparing the power needed to transport the sediment along a particular reach with the power actually available, *critical power* was defined as the power just sufficient to transport the sediment through the reach (Bull 1979, 1991). Interest in the significance of large flood events has led to the estimation of *flood power*, including that related to palaeofloods (Baker and Costa 1987), and a threshold for catastrophic modification of the channel or fluvial landscape has been suggested as a unit stream power of $300\,W m^{-2}$ (Magilligan 1992).

Variations in unit stream power along a river channel have been used to explain patterns of the pool riffle sequence; to determine bedform type for specific sediment size; to relate to the channel HYDRAULIC GEOMETRY; and to explain river channel patterns. Such patterns have been classified according to amount and size of bedload and stream power (Schumm 1981) and channel SINUOSITY has been related to stream power (Schumm 1977). Three major types of floodplain (Nanson and Croke 1992) have been differentiated according to stream power values, including High energy ($\omega > 300\,Wm^{-2}$), Medium energy ($10 < \omega < 300\,Wm^{-2}$), and Low energy ($\omega < 10\,Wm^{-2}$), and stream power variations have also been related to the pattern of the river long profile. These are all ways in which stream power can be related to aspects of *channel morphology*, showing how knowledge of spatial variations of stream power can be the basis for useful applications. In the case of British rivers, Ferguson (1981) demonstrated a thousand fold range in the values of specific power with a clear distinction between values of 100 and $1{,}000\,Wm^{-2}$ in the high runoff, steep slope areas of the west, contrasting with the values between 1 and $10\,Wm^{-2}$ in the low slope, low runoff areas of the south and east. Relations with channel morphology and their spatial variations can be employed to provide explanations for the pattern of long profiles or channel patterns, for example, by considering downstream variations in power to be with minimum unit stream power expenditure, or between equalizing power expenditure and minimizing power expenditure.

Variations in power can also be useful in *river management* problems, as employed by analysing channel adjustments downstream from river channelization works (Brookes 1987); in this case the relationship between bankfull discharge per unit width and water slope was subdivided according to lines of equal specific stream power, showing that eroded sites had specific powers in the range $25–500\,Wm^{-2}$ whereas the remaining sites without change had specific powers between 1 and $35\,Wm^{-2}$. Such examples can be the basis of general guidelines for river managers, with stream power per unit bed area as a criterion for stability in stream restoration projects; a simple classification for guiding river restoration has been developed (Brookes and Sear 1996) by using the ratio between available stream power and erodibility of the substrate. In Denmark, straightened channels tend to recover naturally above a threshold stream power of $35\,Wm^{-2}$ and it is only channels with very high energies that regain some or all of their original sinuosity, so that thresholds could easily be developed for other river environments. Types of river channel adjustment have therefore been related to thresholds of stream power (Brookes 1990), indicating how temporal variations of stream power can be used to understand *variations over time* that have occurred. In a study of the arroyo systems of the northern part of the Henry Mountains in south central Utah (Graf 1983) it was shown that, whereas stream power decreased in the downstream direction during a deposition period which occurred before 1896 when channels were small and meandering, after 1896 total stream power increased in the downstream direction because channels were in the floors of arroyos that confined discharges and resulted in channel erosion and throughput of sediment. In 1980, however, the rate of downstream change in total power was intermediate between the depositional conditions of the 1890s and the erosional conditions of 1909, with deposition occurring in the smallest and largest channels but not in the mid-basin areas. Stream power can also be used as a unifying theme for analyses of urban fluvial geomorphology (Rhoads 1994) and although values are not always easy to calculate it remains a very important variable in fluvial geomorphology.

References

Allen, J.R.L. (1977) Changeable rivers: some aspects of their mechanics and sedimentation. in K.J. Gregory (ed.) *River Channel Changes*, 15–45, Chichester: Wiley.

Bagnold, R.A. (1960) Sediment discharge and stream power: a preliminary announcement, *US Geological Survey Circular* 421.
——(1977) Bedload transport by natural rivers, *Water Resources Research* 13, 303–312.
Baker, V.R. and Costa, J.E. (1987) Flood power, in L. Mayer and D. Nash (eds) *Catastrophic Flooding*, 1–21, Boston: Allen and Unwin.
Brookes, A. (1987) River channel adjustments downstream from channelization works in England and Wales, *Earth Surface Processes and Landforms* 12, 337–351.
——(1990) Restoration and enhancement of engineered river channels: some European experiences, *Regulated Rivers: Research and Management* 5, 45–56.
Brookes, A. and Sear, D.A. (1996) Geomorphological principles for restoring channels, in A. Brookes and F.D. Shields (eds) *River Channel Restoration. Guiding Principles for Sustainable Projects*, 75–101, Chichester: Wiley.
Bull, W.B. (1979) Threshold of critical power in streams, *Geological Society of America Bulletin* 90, 453–464.
——(1991) *Geomorphic Responses to Climatic Change*, Oxford: Oxford University Press.
Ferguson, R.I. (1981) Channel forms and channel changes, in J. Lewin (ed.) *British Rivers*, 90–125, London: George Allen and Unwin.
Graf, W.L. (1983) Downstream changes in stream power in the Henry Mountains, Utah, *Annals of the Association of American Geographers* 73, 373–387.
Gregory, K.J. (1987) The power of nature-energetics in physical geography, in K.J. Gregory (ed.) *Energetics of Physical Environment. Energetic Approaches to Physical Geography*, 1–31, Chichester: Wiley.
Magilligan, F.J. (1992) Thresholds and the spatial variability of stream power during extreme floods, *Geomorphology* 5, 373–390.
Nanson, G.C. and Croke, J.C. (1992) A genetic classification of floodplains, *Geomorphology* 4, 459–486.
Rhoads, B.W. (1994) Stream power: a unifying theme for urban fluvial geomorphology, in E.E. Herricks (ed.) *Urban Runoff and Receiving Systems: An Interdisciplinary Analysis of Impact, Monitoring and Management*, Proceedings of the Engineering Foundation Conference, Mt Crested Butte, Colorado, 4–9 August, 1991, 84, 91–101.
Schumm, S.A. (1977) *The Fluvial System*, New York: Wiley.
——(1981) Evolution and response of the fluvial system, sedimentologic implications, *Society of Economic Palaeontologists and Mineralogists Special Publication* 31, 19–29.

KENNETH GREGORY

STREAM RESTORATION

Stream restoration is the changing of physical, chemical and biological characteristics of a lotic system to match those of a former natural stream condition, or one that has not been disturbed by humans. One common definition of stream restoration is the 'reestablishment of the structure and function of a stream ecosystem' (National Research Council 1992: 17). Ecological restoration, in general, assists the recovery of an ecosystem that has been degraded, damaged or destroyed and restores its historical trajectory (SER 2002).

Other terms in use include stream rehabilitation, reclamation, reconstruction, mitigation, and 'creation' of new functions and values. Reconstructed channels include a new ecological structure so that the desired flora and fauna can return. Rehabilitated or reclaimed streams are partially restored in that only selected functions and values are returned, primarily to serve a human purpose (e.g. flood control, water supply, land stabilization). Most stream restoration projects are actually partial rehabilitations. It is exceedingly difficult to restore all functions and values of the original stream.

Restoring streams that have been routed through pipes is known as 'daylighting'. Many urban (see URBAN GEOMORPHOLOGY) streams have been paved or lined with large rocks, to carry more water faster and without erosion. Efforts to restore these streams are complex, involve many people, lengthy planning times and are very costly (Plate 131).

Plate 131 Concrete-lined channels and restrictive boundary conditions complicate urban stream restorations

Planning: restore to what conditions?

Stream restoration efforts usually cannot achieve pristine or prehistoric conditions because of the massive cultural changes that would be required. Many streams have been dammed, leveed or channelized to convey floodwaters. Full 'restoration' of these streams would require returning processes like flooding, meander migration, channel avulsion, formation and destruction of LARGE WOODY DEBRIS jams, and backwater sedimentation.

Where management approaches alone cannot achieve the desired ecological functions and values, channel reconstructions using large equipment may be needed. With continued monitoring and adaptive management, reconstructed channels can return some lost ecological functions and values.

Streams that function within natural ranges of flow, sediment movement, temperature, channel migration, and other variables, are said to be in dynamic equilibrium. Restoration projects attempt to restore and maintain DYNAMIC EQUILIBRIUM and ecological integrity. Streambank erosion may be part of this dynamic equilibrium if it is balanced and overall dimensions of the channel remain the same. Successful stream restorations integrate geomorphic processes into their designs. Measurement of channel-forming flow conditions is critical to restoring a stable form, pattern and profile (Rosgen 1996).

Restoration design

Design approaches may be based on stream classification and regional curves for hydraulic geometry (Riley 1998; Rosgen 1996), regime and tractive force equations, and reference reaches (Newbury and Gaboury 1993; Rosgen 1996), but the most rigorous approaches feature more sophisticated hydraulic engineering. Risk of failure can be minimized by incorporating the appropriate levels of management (removing disturbance factors) and engineering controls (e.g. weirs, riprap). Planning and design are best accomplished with an interdisciplinary approach (USDA *et al.* 1998), involving knowledgeable fluvial scientists, ecologists and land users.

'Natural channel design' uses reference reaches in dynamic equilibrium with desired ecological functions and values. Stream classification systems have been developed to assist this process and to promote accurate stream morphological descriptions (Rosgen 1996). When used with experience and scientific design approaches, natural channel designs can result in successful restorations. When used as a 'carbon copy' or cookie-cutter approach, however, the results can be less than successful.

Most stream restoration projects involve modifying an existing stream's location, alignment, meander pattern (see MEANDERING) cross-section dimension, longitudinal profile, or aquatic or terrestrial habitat. Streambank erosion, sediment transport, flooding and sediment deposition are processes that support ecological functions. Having these physical processes become self-sustaining is what separates stream restoration from stream stabilization, stream reconstruction or stream rehabilitation projects.

Soil bioengineering

Soil bioengineering can be an important stream restoration component (Figure 159). Live plants, plant materials and man-made materials are used as systems that interface with Earth materials to create stable streams and banks, and achieve the desired ecological functions and values. Practical application of soil bioengineering techniques requires knowledge of overall performance criteria, design flow and sediment conditions, and habitat suitability.

References

National Research Council (1992) *Restoration of Aquatic Ecosystems: Science, Technology, and Public Policy*, Washington, DC: National Academy Press.
Newbury, Robert W. and Marc N. Gaboury (1993) *Stream Analysis and Fish Habitat Design, A Field Manual*, Gibsons, British Columbia Newbury Hydraulics Ltd, Canada.
Riley, Ann L. (1998) *Restoring Streams in Cities: A Guide for Planners, Policymakers, and Citizens*, Washington, DC: Island Press.
Rosgen, Dave (1996) *Applied River Morphology*, Wildland Hydrology, Minneapolis, MN: Printed Media Companies.
SER (2002) *The SER Primer on Ecological Restoration*, Society for Ecological Restoration Science and Policy Working Group, www.ser.org/
USDA et. al. (10/1998) *Stream Corridor Restoration: Principles, Processes, and Practices*, the Federal Interagency Stream Restoration Working Group, www.usda.gov/stream_restoration

Further reading

Brookes, A. and Shields, F.D. Jr. (eds) (1996) *River Channel Restoration: Guiding Principles for Sustainable Projects*, Chichester: Wiley.

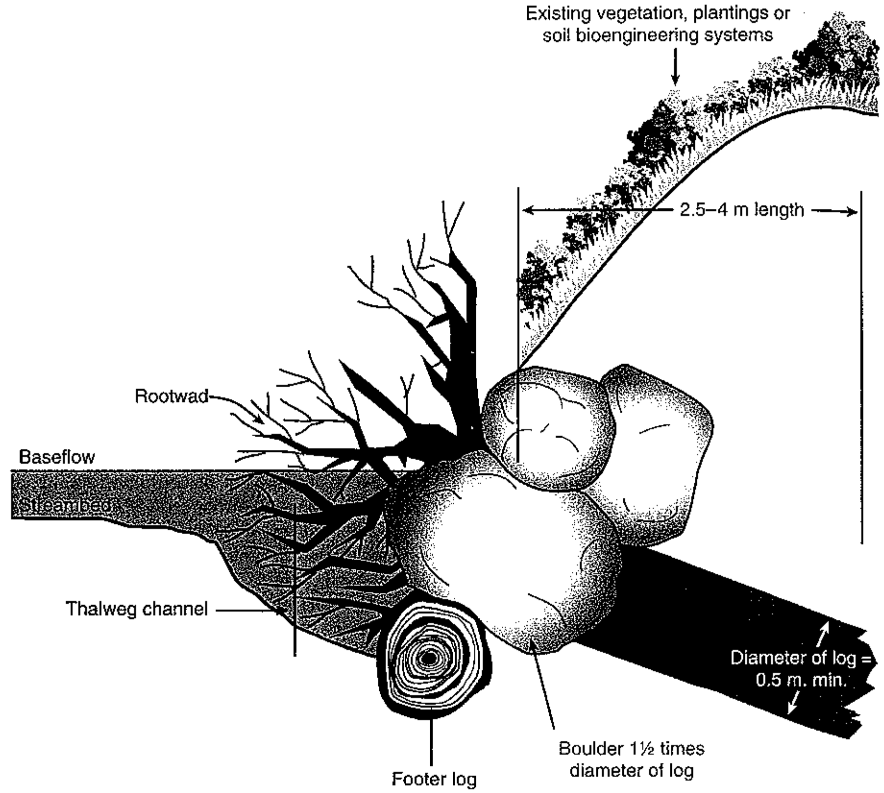


Figure 159 Soil bioengineering example: rootwad

Dunne, T. and Leopold, L.B. (1978) *Water in Environmental Planning*, New York: W. H. Freeman.

Lane, E.W. (1955) The importance of fluvial geomorphology in hydraulic engineering, *American Society of Civil Engineering, Proceedings* 81, paper 745, 1–17.

Leopold, L.B. (1994) *A View of the River*, Cambridge: Harvard University Press.

Raine, A.W. and Gardiner, J.N. (1995) *Rivercare: Guidelines for Ecologically Sustainable Management of Rivers and Riparian Vegetation*, LWRRDC Occasional Paper Series No. 03/95, Canberra: Land and Water Resources Research and Development Corporation.

Ward, D. and Holmes, N.; Paul José (eds) (1995) *The New Rivers and Wildlife Handbook*, National Rivers Authority, Royal Society for the Protection of Birds, and the Royal Society for Nature Conservation: The Lodge, Sandy, Bedfordshire, UK.

SEE ALSO: bankfull discharge; biogeomorphology; fluvial geomorphology; mining impacts on rivers; riparian geomorphology; river continuum; river restoration; sediment budget; step-pool system; stream ordering; stream power

JERRY M. BERNARD

STRIATION

Striations are shallow scratches or grooves cut by brittle impact into rock surfaces, boulders or pebbles. Striations may be up to a metre or more in length. They occur widely in areas of former glacial erosion where rock fragments, sand and silt grains transported in the basal ice have impacted surfaces as the ice moved forward jerkily by basal sliding. Some striations have nail head or wedge shapes, with the broad section being at the down-ice end. Striations occur frequently in glacierized areas, especially on fine-grained physically hard rocks such as quartzites and massive limestones. Glacial polish often occurs with striae. Crossing striations may reflect local ice flow variations, changes in ice flows from different glacial centres during one glaciation, or multiple stages of glaciation. Striations may also form where sea-ice, lake-ice, snow banks, screes, large rock masses and debris creep, flow, avalanche, slide, fall or shear over or onto rock surfaces. ICEBERGS may striate the upper surfaces of clasts in current-winnowed boulder lag horizons of glacimarine deposits. Detritus carried by high velocity water flow in jökulhlaups can also form striations in channels eroded in rock.

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.

ERIC A. COLHOUN

STROMATOLITE (STROMATOLITH)

A term first used by Kalkowsky (1908) to describe some sedimentary structures in the Bunter of North Germany. A more modern definition (Walter 1976: 1) is that they are 'organosedimentary structures produced by sediment trapping, binding and/or precipitation as a result of the growth and metabolic activity of micro-organisms, principally cyanophytes'. They can develop in marine, marsh and lacustrine environments and, though they form today where conditions permit, they reached the acme of their development in the Proterozoic (Hofman 1973). The largest known forms are mounds several hundreds of metres across and several tens of metres high. Gross morphologies vary in the extreme and range from stratiform crustose forms, through nodular and bulbous mounds and spherical oncoids, to long slender columns, erect to inclined, and with various styles of branching. Classic examples are known from coastal regions like those at Shark Bay in Western Australia and from pluvial lake shorelines in areas like the Altiplano of Bolivia where they form massive calcareous encrustrations and bioherms (Rouchy *et al.* 1996).

References

Hofman, H.J. (1973) Stromatolites: characteristics and utility, *Earth-science Reviews* 9, 339–373.

Kalkowsky, E. (1908) Oolith und Stromatolith im norddeutschen Buntsandstein, *Zeitschrift Deutsche Geologische Gesselschaft* 60, 68–125.

Rouchy, J.M., Servant, M., Fournier, M. and Causse, C. (1996) Extensive carbonate algal bioherms in upper Pleistocene saline lakes of the central Altiplano of Bolivia, *Sedimentology* 43, 973–993.

Walter, M.R. (1976) Stromatolites, *Developments in Sedimentology* 20.

A.S. GOUDIE

STRUCTURAL LANDFORM

Structural landforms are those which in their appearance reflect, and are adjusted to, geological structure of underlying bedrock. This effect is achieved through direct or indirect control which structural elements exert on the course and intensity of exogenic processes shaping the landforms. 'Structure' here is usually understood *sensu lato*, i.e. it encompasses such diverse phenomena as facies differentiation, lithological contrasts, fracture patterns, faults and folds, tectonic disposition of strata, geometries of intrusive and extrusive bodies, etc. In general, structural landforms develop through differential weathering and erosion which exploit the structures and emphasize unequal resistance of adjacent rock complexes, whereas relief features created by direct action of endogenic forces fall into the category of tectonic landform (see TECTONIC GEOMORPHOLOGY).

Following the above definition, structural landforms may be subdivided into several categories. There are landform assemblages specific for certain rock types, for example granite (see GRANITE GEOMORPHOLOGY) or carbonate rocks (see KARST). In many cases, their development and appearance are controlled by jointing patterns, in the way that joints of regional extent (master joints) and zones of dense fracturing are exploited towards topographic depressions, whereas more massive compartments are left standing as residual hills, uplands or big boulders. Thus, joint-aligned valleys, rows of sinkholes, basins at joint intersections, and many TORs and domed INSELBERGs may be regarded as structural landforms. A number of small-scale features, such as KARREN on carbonate rock outcrops or deep clefts, may develop along joints and are therefore also structural. In igneous rocks in particular, many structural features originating at the consolidation stage subsequently become avenues for weathering and become decisive for the shape of minor and medium-scale landforms (Twidale and Vidal-Romani 1994).

Another group includes landforms reflecting the variable dip of sedimentary strata and, at the same time, the unequal resistance of consecutive layers against exogenic agents. Undeformed, horizontal layers give rise to plains, if at low altitude, or plateaux, if elevated and bounded by marginal escarpments. A plain or a plateau surface is usually underlain by a resistant rock layer, such as quartz sandstone or massive limestone. Dissection of a plateau may expose underlying strata of variable resistance, the stronger of which will support structural benches on valley sides, such as in the Grand Canyon of the Colorado River. Tilting of strata induces differential denudation, in the course of which rock complexes of lower strength are eroded into valleys or rolling plains, whereas more resistant rocks give rise to parallel ridges, escarpments or mid-slope ledges. If the dip is less than 10°, highly asymmetric ridges called CUESTAs develop. With the dip in the range 10°–30°, less asymmetric monoclinal (or homoclinal) ridges form, whereas in the case of even steeper tilt a symmetric ridge named a HOGBACK will originate. In areas built of dipping sedimentary rocks, drainage patterns usually show much adjustment to structure too. Dendritic patterns are typical for negligible dip, whereas with increasing differential erosion they will evolve into trellis patterns. The spatial arrangement of ridges depends on the regional pattern of deformation. If a simple tilt is involved, ridge axes will generally follow straight lines, perpendicular to the dip. If a central part of a former sedimentary basin is downwarped, concentric ridges facing outwards will develop (the Paris Basin is an example), whereas inward facing ridges will typify domed structures with breached central parts.

In mountain areas built of folded sedimentary rocks typical structural landforms are anticlinal ridges and synclinal valleys or, in the case of INVERTED RELIEF, anticlinal valleys and synclinal ridges. Asymmetric homoclinal ridges and hogbacks occur frequently too, but the degree of structural deformation in mountains is usually so high that the spatial extent of these landforms is limited.

A further category includes landforms built of igneous rocks, the present-day appearance of which may reflect the way of magma emplacement. Large granite intrusions in post-orogenic settings may assume the form of large-radius domes (laccoliths) which, after unroofing, are reflected in the topography as upland terrains, sloping in all directions. Dartmoor upland in south-west Britain is one example. Rising igneous domes induce updoming of overlying strata which then may be differentially eroded towards triangular faces called FLAT IRONs (Ollier and Pain 1981). Smaller linear intrusions, i.e. dykes and sills, are typically composed of material more resistant than the host rock, hence differential denudation leaves them appearing as laterally extensive, vertical, jagged ridges (for dykes) or topographic steps (for sills). Denudation of a former volcano may expose deeper parts of the vent filled with solidified, resistant magma, which then becomes a steep-sided conical hill, called a neck. Necks often display impressive columnar jointing of rock, the most famous example being perhaps Devil's Tower in Wyoming, USA.

Many landforms built of extrusive rocks may also be regarded as structural, including rhyolitic domes and plugs, sloping flanks of shield volcanoes, or plateaux underlain by horizontal or gently inclined lava flows (traps). Subsequent tilting and erosion of a multiple lava flow area may produce an assemblage of landforms similar to those developed on tilted sedimentary rocks.

Structural landforms are of various sizes, from features of regional extent to local manifestations of small-scale structures. Mega-scale examples are extensive plains developed upon flat-lying strata or dome mountains. Medium-scale landforms include plateaux, cuesta ridges, domed hills and intermontane basins. Even smaller are tors and mid-slope benches, whereas joint-aligned pools in a bedrock river bed would indicate the most localized structural control.

References

Ollier, C.D. and Pain, C.F. (1981) Active gneiss domes in Papua New Guinea: new tectonic landforms, *Zeitschrift für Geomorphologie* 25, 133–145.
Twidale, C.R. and Vidal-Romani, J.R. (1994) On the multistage development of etch forms, *Geomorphology* 11, 157–186.

Further reading

Gerrard, J. (1986) *Rocks and Landforms*, London: Unwin Hyman.
Peulvast, J.-P. and Vanney, J.-R. (2001) *Géomorphologie structurale – Terre, corps planétaires solides, tome 1*, 2, Paris: Gordon and Breach
Yatsu, E. (1966) *Rock Control in Geomorphology*, Tokyo: Sozosha.

PIOTR MIGOŃ

STURZSTROM

A sturzstrom is a high volume of mostly dry rock material caused by the collapse of a slope or cliff created by large falls and slides moving at high velocities and for long distances, even on a gentle slope (Hsü 1975). Sturzstroms can reach velocities of over $50\,\mathrm{m\,s^{-1}}$ and can travel over distances of kilometres. The accumulation volume may exceed 1 million m^3, covering a total surface of over $0.1\,\mathrm{km}^2$. In relation to its velocity and dimensions, this kind of landslide can be extremely costly in terms of human lives and damage.

Alternative terms for sturzstrom are rock avalanche, rockfall avalanche or rock-slide avalanche (Angeli *et al.* 1996). Examples of historic events are the Elm sturzstrom of 1881 in Switzerland (Heim 1932), the Valpola rock avalanche of 1987 in the Italian Alps and the Frank landslide of 1903 in Canada. In Europe the highest concentration of these phenomena is found in the Northern and Southern Calcareous Alps (Abele 1974).

A sturzstrom can develop (1) by the fall or slide of a rock body which during movement progressively looses its cohesion by turning into dry debris and, thus, continues its advancement as a debris avalanche, (2) by the sudden mobilization of a debris deposit by a debris avalanche or DEBRIS FLOW, either because of the fall of an overhanging rock mass or by seismic shocks.

Although several investigations have been carried out, no universally accepted explanation has yet been proposed. The mechanical analysis of sturzstroms includes two stages: the initial failure and subsequent streaming. Explanations include turbulent grain flow with dispersive stresses arising from momentum transfer between colliding grains (Cruden and Varnes 1996), fluidization of particles caused by incorporating air, the existence of a cushion of trapped air, acoustic fluidization or rock melting by frictional heat (Erismann and Abele 2001).

In alpine regions the ice retreat of the last alpine glaciation caused significant stress relief on mountain slopes. This unloading is one of the causes of Holocene sturzstroms. However, there is evidence that the process not only took place shortly after the ice retreat (as found between 12,000 and 10,000 BP) but that the loss of shear strength needed much more time to cause the process. The Eibsee sturzstrom (Zugspitze, German Alps, volume: $400 \times 10^6\,\mathrm{m}^3$) has been dated at 3,700 years BP which may fit this hypothesis. The exact triggering causes which may enable realistic predictions of sturzstroms requires further attention.

Sturzstroms are highly destructive. Once an event has occurred it is important to discover whether the mass has dammed the valley. If a lake has formed, maximum effort must be given to take control of the breaching dam, because the resulting flow may cause a second disaster.

References

Abele, G. (1974) *Bergstürze in den Alpen*, Wissenschaftl, Alpenvereinshefte, 25, München.
Angeli, M.G., Gasparetto, P., Menotti, R.M., Pasuto, A., Silvano, S. and Soldati, M. (1996) Rock avalanche, in R. Dikau, D. Brunsden, L. Schrott and M.-L. Ibsen (eds) *Landslide Recognition*, 190–201, Chichester: Wiley.
Cruden, D.M. and Varnes, D.J. (1996) Landslide types and processes, in A.K. Turner and R.L. Schuster (eds) *Landslides. Investigation and Mitigation*, 36–75, Washington: National Academy Press.
Erismann, T.H. and Abele, G. (2001) *Dynamics of Rockslides and Rockfalls*, Heidelberg: Springer.
Heim, A. (1932) *Bergsturz und Menschenleben*, Zürich: Fretz und Wasmuth (English trans. N. Skermer, (1989) *Landslides and Human Lives.*, Vancouver, Canada: BiTech Publishers.
Hsü, K.J. (1975) Catastrophic debris streams (sturzstroms) generated by rockfall, *Geological Society of America Bulletin* 86, 129–140.

RICHARD DIKAU

SUBAERIAL

Refers to all conditions or processes occurring in the open air or on the land surface (e.g. subaerial weathering) as opposed to those occurring in submarine (underwater) or subterranean (underground) environments. The term also applies to materials and features that are created and/or located on the Earth's surface (e.g. subaerial aeolian dunes, subaerial volcano, etc.), and sometimes is inclusive of fluvial forms (of rivers). The notion that all things in the landscape are created by subaerial processes and conditions is termed subaerialism.

STEVE WARD

SUBCUTANEOUS FLOW

The lateral transfer of water in the subcutaneous (or epikarstic) zone. The subcutaneous zone is a highly weathered region in well-developed KARST environments lying in the upper part of the percolation zone, between the soil and the relatively unweathered and permeably saturated phreatic zone below. When water stored in the subcutaneous zone is full (e.g following precipitation) a potentiometric surface (epikarstic water table) is produced. Any further input is subsequently transferred laterally, with movements of water occurring along preferred pathways (neighbouring joints and shafts) down the hydraulic gradient in the epikarstic water table (Williams 1983). Corrosion is intensified where flow routes converge above the more efficient paths, resulting in differential surface lowering, accentuating over time and by irregular distribution of solution. Uniform percolation through the subcutaneous zone will result in uniform surface lowering. Rates of subcutaneous flow vary considerably, commonly between 2–10 weeks, but are hard to gauge (especially in old, well-developed karsts). The term is also used to refer to the process of piping in soils.

Reference

Williams, P.W. (1983) The role of the subcutaneous zone in karst hydrology, *Journal of Hydrology* 61, 45–67.

Further reading

Gunn, J. (1981) Hydrological processes in karst depressions, *Zeitschrift für Geomorphologie* 25(3), 313–331.

SEE ALSO: epikarst

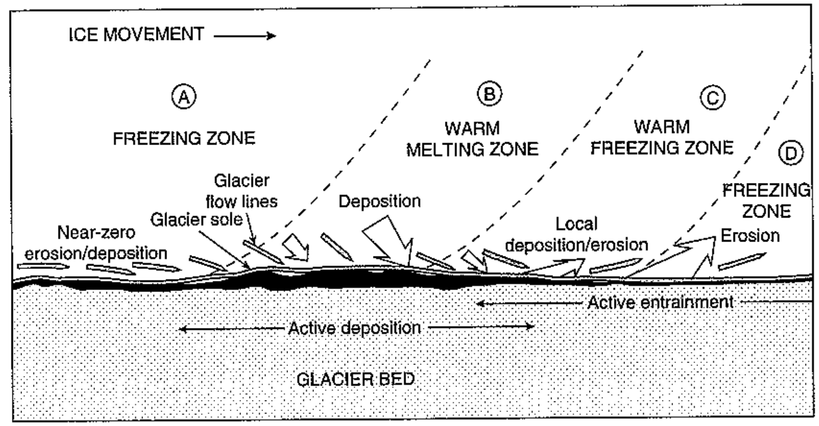
STEVE WARD

SUBGLACIAL GEOMORPHOLOGY

Subglacial geomorphology has the most profound and indelible imprint on any glaciated landscape. Subglacial geomorphic processes are among the most complex, yet least understood set of glacial processes. Our restricted knowledge stems from the inaccessible nature of what occurs beneath an ice mass and the limited extent of modern analogues. No other aspect of glacial geomorphology impacts on the daily lives of millions of people to such a degree as does subglacial geomorphology, for example, in terms of foundations, roads, railroads, waste disposal sites, agriculture, aquifers and construction materials.

The subglacial environment is that glacial subsystem directly beneath an ice mass that includes cavities and channels that are not influenced by subaerial processes. Subglacial geomorphology considers all aspects of topographic change beneath ice masses as a result of erosional and depositional processes (Figure 160). The subglacial environment is a boundary interface where complex sets of processes interact altering the morphological, thermal and rheological states of the interface between the ice mass and its bed. The key to understanding subglacial geomorphology lies in the mechanics of this interface. This boundary zone migrates across the landscape with every ice advance and retreat. All terrains covered by glaciers have been affected and altered by the passage of this interface. This interface is a function of the prevailing basal ice and bed conditions; a complex relationship between basal ice dynamics, sediments and bedrock, subglacial hydraulics and the glacier bed ambient temperature. Changes in basal ice and/or bed conditions may be widespread or local, and develop rapidly or slowly. These fluctuating conditions may be of enormous magnitude or simply be minor variations at the interface, the former being detectable whilst the latter may leave little or no imprint on subglacial geomorphology.

Subglacial geomorphological processes are constrained by the temperature at the ice–bed interface. Basal thermal regimes can be either temperate or polar but in all likelihood most ice masses are polythermal. In temperate, wet-based glaciers, typical of almost all ice masses today and most ice masses during the Pleistocene and earlier, basal temperatures are found at −1 to −3 °C. In polar, cold, dry-based glaciers, typical of a few isolated central areas of East Antarctica today and possibly the central areas of the vast Pleistocene ice sheets, the basal parts of an ice mass are frozen to the bed with temperatures of −13 to −18 °C (Van der Veen 1999). Thermal conditions at the ice–bed interface are more



Ⓐ Glacier totally frozen to its bed

Ⓑ Glacier melting at bed with isolated patches remaining frozen, larger

Ⓒ Glacier beginning to freeze to bed, patches melting toward larger

Figure 160 Models of subglacial thermal regimes, their spatial relationships and processes of subglacial erosion, transportation and deposition

complex than implied by these two thermal states, in fact basal ice temperatures vary temporally and spatially producing polythermal conditions (Menzies and Shilts 2002).

In considering ice–bed interface temperature fluctuations, the following parameters converge to establish the specific thermal conditions: (a) rate of snow accumulation and snow temperature; (b) geothermal heat flux; (c) mean annual surface air temperature; (d) ice surface velocity; (e) basal ice velocity; (f) subglacial meltwater flux and temperature; and (g) the imprinted 'memory' of previous thermal conditions (Figure 161a). The relevance of these two extreme thermal states is that under temperate conditions, meltwater occurs at the ice–bed interface, basal ice containing debris can be released, and meltwater processes and/or saturated debris moving as a deforming layer can be accomplished thereby permitting various landforms to develop. Under polar conditions, with the ice mass frozen to its bed, no meltwater is present and only ice movement by plastic deformation occurs thereby geomorphological processes are constrained but not altogether suspended.

It is apparent that polythermal bed conditions prevail in the long term under any ice mass, the switch from wet to dry and back again repeatedly results in localized erosion and deposition of much glacial sediment. Although long distance transport does occur beneath and within ice masses the dominant form of transport is short distance of <10 km resulting in subglacial sediments being typically locally derived, transported and deposited. Figure 160 illustrates the probable relationship between areas of dominantly erosive glacial action and depositional processes under ice sheet conditions. Beneath valley glaciers, due to much shorter distances and thinner ice cover (lower basal ice pressures), a similar, but less complete, set of conditions may prevail (Benn and Evans 1998).

Subglacial erosional landforms

Subglacial erosion processes are pervasive at the active ice–bed interface. In some terrains under certain subglacial conditions, erosive processes become dominant. Almost any description of the impact on the land surface of past glaciers focuses on the grandeur and size of FJORDS and the sculpted bedrock features of once glaciated terrains. However, our understanding of how these distinctive features were fashioned remains limited. Erosional forms exist at an immense range of

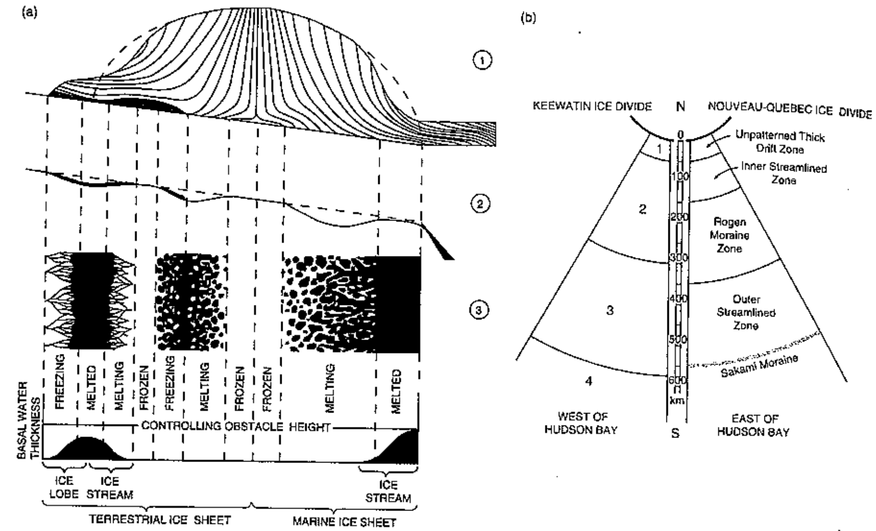


Figure 161 Subglacial thermal zonation in relation to subglacial processes beneath a steady-state ice sheet with terrestrial and marine margins. 1, internal ice flow trajectories; 2, subglacial topography created by subglacial erosion; 3, areal distribution of subglacial meltwater (reprinted with modification after Hughes 1995)

scales from surface microscopic features on particles to the megascale landforms. Regional erosional features can be subdivided into regional and local forms. However, some forms, such as ROCHES MOUTONÉES, occur at all scales. Regional erosional features are either areal (spatially pervasive) or linear (spatially discrete) landscape types. In the former case, scoured terrain develops where limited debris existed at the ice–bed interface and dominantly polar bed conditions prevailed. In contrast, linear erosion processes are differential in their impact upon terrain being confined within specific areas. Linear erosional forms are indicative of subglacial bed states in which rapid but spatially restricted basal ice movement and/or meltwater channelling has occurred.

Terrains of regional areal erosion typically exhibit low relief amplitude, limited sediment deposition and have a moulded and scoured appearance. Geological structure has often been partially exhumed in these terrains and irregular depressions and small roches moutonnées occur. These terrains, known as 'knock and lochan' in north-west Scotland, are typical of this form of GLACIAL EROSION. Similar landscapes exist in shield terrains in Canada (Plate 132a) and Fennoscandia, along the edges of the Greenland and Antarctic Ice Sheets, in Patagonia, and South Island, New Zealand. The form of this regional erosional landscape appears related to relatively slow-moving ice masses under polar-bed conditions with limited debris present. A range of wear processes operate across bedrock surfaces where occasional protuberances lead to ice pressure melting and meltwater discharges often under high pressure. Within this landscape, at a lower scale, crag and tails, and roches moutonnées are prevalent. Typically, P-forms and associated forms related to rapid subglacial meltwater flow are common.

Beneath specific zones within ice sheets, regional linear erosion appears to occur probably as a consequence of ice streaming and fast-moving, but spatially restricted, basal ice. Under these conditions, major linear forms of glacial erosion develop, such as incised bedrock troughs, fjords and tunnel valleys. At meso and microscale fluted bedrock, striae and grooves are witness to this style of selective erosion.

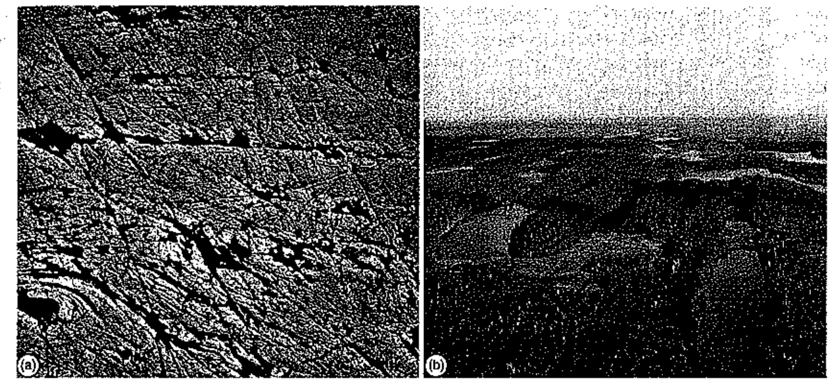


Plate 132 (a) Areally scoured glaciated landscape of the Canadian Shield near La Troie, Quebec (photo courtesy Government of Canada); (b) subglacial depositional landscape of drumlins of the Kuusamo drumlin field, Finland (photo courtesy of R. Aario)

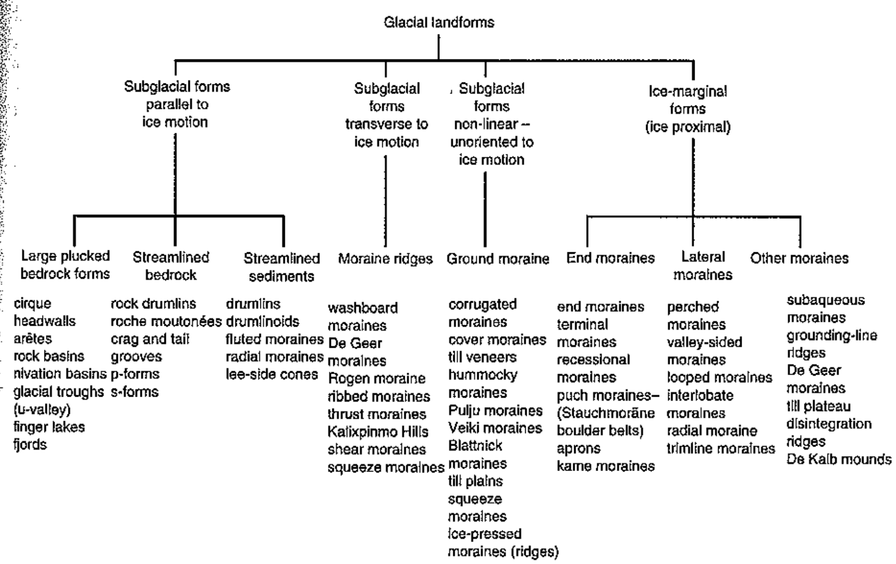


Figure 162 Subglacial landform types

Subglacial depositional landforms

Subglacial landforms can be subdivided into those developed under (1) active ice flow (advance and active retreat phases); and (2) passive or 'dying' ice flow (retreat phase). It is apparent that at least one group of landforms, previously considered as separate entities, may be related to each other as variant bedforms in a continuum of landforms at the subglacial interface. Thus drumlins, fluted moraine and Rogen moraine may be a spectrum of forms subglacially developed that evolve as a function of thermal, topgraphic, glaciological and rheological conditions (Plate 132b).

Suites of landforms can be attributed to active ice and to indirect or passive ice action (Figure 162). Most of the forms are constructional under active and/or indirect passive ice action while others may be partially erosional following ice retreat and the influence of subaerial processes. The location and type of subglacial landforms reflect the complex patterns of processes beneath the ice in differing locations under varying glacidynamic conditions (Figure 161b).

Subglacial geomorphology is a reflection of a myriad of erosional, transportational and depositional processes that produce a complex and varied glaciated landscape within which the influence of underlying topography may be muted due to the effects of deposition or sharpened due to erosional processes. The uniqueness, for example, of fjord landscapes or the gently rolling plains of the Prairies in North America stand as remarkable testaments to the impact on the Earth's surface of subglacial geomorphology.

References

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.

Hughes, T.J. (1995) Ice sheet modelling and the reconstruction of former ice sheets from glacial geo(morpho)logical field data, in J. Menzies (ed.) *Modern Glacial Environments*, 77–100, Oxford: Butterworth-Heinemann.

Menzies, J. and Shilts, W.W. (2002) Subglacial environments, in J. Menzies (ed.) *Modern and Past Glacial Environments*, 183–278, Oxford: Butterworth-Heinemann.

Van der Veen, C.J. (1999) *Fundamentals of Glacier Dynamics*, Rotterdam: A.A. Balkema.

JOHN MENZIES

SUBMARINE LANDSLIDE GEOMORPHOLOGY

The term submarine landslide is often used as a generic descriptor for an erosive and/or depositional feature, of sediment or rock, preserved on the seafloor, or below it (if ancient) after a MASS MOVEMENT. Less generic definitions of seafloor failures such as DEBRIS FLOWS, avalanches, spreads, and the like suggest a rheology of the material as it fails. This is often hard to discern when only a scar is left as evidence of erosion, and as evidence of deposition only the rare sediment or debris pile that can be discerned by remote sensing techniques. An excellent review of submarine landslides (Plate 133) can be found in Hampton *et al.* (1996).

Detection

Most submarine landslides are old and may take place unnoticed if contemporary. Therefore, evidence of an event is obtained by remote sensing techniques such as seismic reflection, side-scan sonar and multibeam bathymetry. Data are then analysed using a seismic interpretation package that can keep track of individual horizons, or assembled in a Geographic Information System (GIS) database.

Remote sensing techniques used in the ocean employ the reflection of acoustic waves to measure the depth and physical properties of material on the surface and in the subsurface. In general, higher frequency acoustic energy yields higher resolution returns, but the energy will not propagate as deep. A 12 kHz (kilohertz) signal is used to measure water depth, and has very little subseafloor penetration, whereas a 3.5 kHz signal can penetrate over 100 m in soft sediment, and a high power, low frequency airgun source can penetrate many kilometres below the seafloor. Once a potential feature is identified, groundtruthing and geotechnical analysis of the material can be done using a variety of coring devices.

Types

Submarine mass movements have many characteristics similar to subaerial landslides, and much of the terminology is the same. Submarine failures tend to be much larger, and often better preserved over the timescale of tens of thousands of years due to slow diffusion processes in deepwater environments. Despite their size and preservation, submarine failures are hard to detect, and there is seldom a witness to describe the dynamics of the slide as it occurs. Furthermore, there is a tendency for submarine sediment failures to disintegrate, leaving a deposit that is difficult to discern on the surface. We are left to make inferences based on the resulting morphology of the erosive and depositional features, and physical properties of the surrounding material.

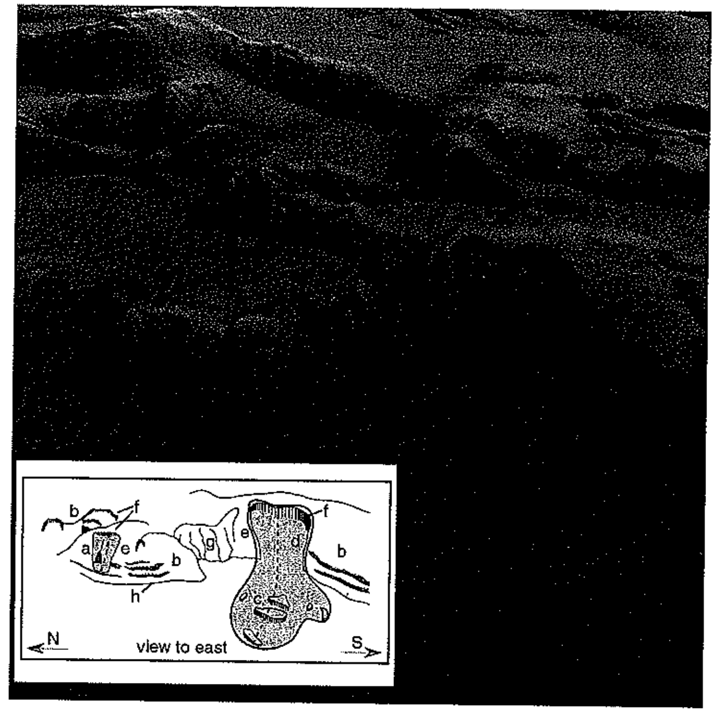


Plate 133 Submarine landslides imaged using multibeam bathymetric data from the base of the convergent continental margin, offshore Oregon, USA (modified from McAdoo *et al.* 2000). Approximately 3× vertical exaggeration, view is looking east with north to the right. The field of view at the base of the slope is approximately 30 km. The letters in the inset line drawing are referred to below. Disintegrative slides (a) have characteristic scar morphology, but lack bathymetric evidence of failed material at the base of the slide. Blocky slides have cohesive blocks of material (c) that can be related to a particular slide scar (d). The unfailed slope adjacent to the failure scar (e) is used as a proxy for the pre-failure slope conditions. Regions of the slope where material has clearly been eroded (g) are difficult to classify as a landslide as they lack characteristic evidence of discrete slide events such as a distinct headscarp (f). A series of curious terraces (h) rim the very base of the slope, and often coincide with the lowermost terminus of the slides, resulting in a 'Hanging Slide' (a). Notice the very steep seafloor with little visible evidence of erosion between slides on the westernmost ridge (closest) and on the other background ridges (b)

Failed material can either remain cohesive (a 'blocky' slide) or disintegrate into a flow. *Mass movements* (translational and rotational slides and slumps) move downslope under the force due to gravity and other body forces. The physics of *mass flows* (TURBIDITY CURRENTS, avalanches, debris flows, etc.) is governed by tractive stresses associated with a fluid.

A displaced mass of sediment or rock originates from and moves along a *rupture surface* (failure or slide surface), and portions of that failed mass may remain in contact with the rupture surface. If the rupture surface is curved (concave upward), the displaced mass rotates, and the failure is often termed a *rotational* slide or slump. If the failure plane is flat, say a bedding plane, then the slide is described as *translational*. As a block of failed material moves downslope it may disintegrate and become either a debris flow (large clasts supported by a fine-grained matrix) or *turbidity* *current* (turbulent flow of fluid supported sediment; Morgenstern 1967). LIQUEFACTION occurs when repacking of sediment grains causes a buildup of fluid pressure to the point where individual grains lose contact with each other, reducing the material's shear strength to nil. This often occurs during earthquakes, and mass flows can occur on very low inclines.

Mechanics

Failure occurs when the downslope-oriented shear stress exceeds the shear strength of the slope material. As stresses for any given environment are likely to be similar (with the possible exception of earthquakes), there is a tendency to focus on the strength side of the equation, which can range from unconsolidated sediment to well-lithified igneous rock.

The strength can be defined by the effective cohesive strength (C_0), the stress acting normal to the failure surface (σ_n), the PORE-WATER PRESSURE (p_f) and the angle of internal friction (ϕ)

$$\tau_f = C_0 + (\sigma_n - p_f) \tan \phi$$

Many submarine landslides occur on slopes well below the angle of repose (see REPOSE, ANGLE OF), therefore require a mechanism to reduce the strength of the sediment at the locus of failure. For a material with a given C_0 and ϕ, the reduction in strength can come from either an increase in p_f or a reduction in σ_n. Sediment strength is not affected by water depth, as hydrostatic conditions prevail to the depths of most slides. Therefore, rock or sediment on the seafloor will not fail on slopes less than ϕ without a transient of some sort.

There are several processes that increase p_f which in turn reduces the EFFECTIVE STRESS ($\sigma_n - p_f$), hence reducing the strength of the material. Gas hydrates (a methane-ice compound found in the sub-seafloor of many continental margins) dissociate with a fall in sea level (reduction of pressure) or an increase in temperature, and release bubble-phase methane which can get trapped, leading to an increase in p_f. Large failures that occurred during sea-level lowstand and may be related to hydrate dissociation include the Storegga slide offshore Norway, where 5,000 km^3 of material failed 8,000 years ago (Bouriak *et al.* 2000), and numerous slides in the Beaufort Sea (Kayen and Lee 1991). Earthquake generated seismic waves propagating through sediment can also increase pore-water pressures as the waves pass rapidly enough to prevent pressure dissipation (Lee and Edwards 1986). Accelerations from earthquakes can cyclically reduce and increase σ_n, therefore the resulting increase in p_f (which is frequency dependent) may be as important as a 'critical acceleration' required for failure. In shallow water, storm waves may also generate increased pore pressures due to cyclic loading, and failure can be initiated in zones of increased shear stress halfway between the crests and troughs of the waves (Seed and Rahman 1978). These failures will occur in shallow water (continental shelf and upper slope), and should not be deep-seated.

Post-failure evolution

Depending on the stresses present, sediment properties, and slope morphology, a failed mass can either stop a short distance from the source, or travel very long distances. One method of determining the likelihood of a sediment failure to disintegrate or remain as cohesive packages under cyclic loading is to consider the pre-failure water content and steady-state effective stress (see Hampton *et al.* 1996). For material with a high water content and effective stress, pore pressure is likely to increase during a transient loading event, and the effective stress will decrease, resulting in a tendency towards disintegrative failures ('contractive' sediment, as the fabric tends to collapse). In contrast, sediment that already has a high effective stress and lower water content, cohesive failures are more likely to occur as the sediment dilates, reducing the effective stress.

The post-failure dynamics are a complex combination of viscous and plastic elements. Norem *et al.* (1990) proposed an approach to modelling submarine landslides with a combination of these plastic and viscous terms:

$$\tau = \tau_c + \sigma_n \{1 - (p_f/\rho_s H)\} \tan \phi + m\rho\, (dv/dy)^r$$

where τ is the total shear strength mobilized during the flow, τ_c is the yield strength (similar to C_0 in the static case), $p_f/\rho_s H$ is the ratio of pore pressure to sediment density (ρ_s) and failure thickness (H), with an additional viscous term where m is the Bingham viscosity, ρ is the fluid density and a vertical velocity gradient term $(dv/dy)^r$ where r ranges between 1 for macroviscous and 2 for inertial flows. This formulation may be able to predict the final shape of a failure deposit, but does not explain the incredible mobility of slides such as the Nuuanu slide off Hawaii, where individual blocks travelled hundreds of kilometres. A hydroplaning model, where the failed material traps a layer of water above the rupture surface, helps explain the mobility of some failures (Mohrig *et al.* 1998)

Location

Submarine landslides are ubiquitous throughout the oceans, occurring on ACTIVE MARGINs and PASSIVE MARGINs, continental shelves and slopes, in the heads of FJORDs and near active deltas, on the seaward side of barrier reefs (see BARRIER AND BARRIER ISLAND), and on the flanks of mid-ocean ridges, seamounts, GUYOTs, or volcanic islands. Triggers of offshore landslides include earthquakes, large storms, SEA LEVEL change, rapid sedimentation, gas hydrate dissociation, or channel erosion in SUBMARINE VALLEYs and CANYONS among other things. Listed below are environments particularly susceptible to failure.

RIVER DELTAS

As regions of rapid sedimentation and erosion in canyons and channels, RIVER DELTAs offer zones of steep slopes and high fluid pressure that facilitate failure. When deposited on the seafloor, fine-grained sediment often has high water content yet low permeabilities, which can lead to fluid overpressure and underconsolidation. When deposited, organic-rich sediment decays producing bubble phase hydrocarbons, which subsequently aids overpressure.

The steady accumulation of fluid overpressures, and the increasing slopes are not sufficient to generate a noticeable failure – a transient such as a storm or earthquake is a necessary trigger. In 1969, large swells from Hurricane Camille in the Gulf of Mexico cyclically loaded the seafloor sediment yielding a build-up of fluid pressure that caused a failure in 100 m of water. In 1980, a magnitude 7.0 earthquake caused liquefaction of overpressured sediment on a 0.25-degree slope in the Klamath River delta, offshore northern California.

FJORDS

The offshore portion of deltas at the heads of FJORDs is a region highly susceptible to landslides. The high, often organic-rich, sediment loads of streams that drain glaciers can create elevated fluid pressures in the delta due to rapid sedimentation and decomposition that creates gas. The combination of fine-grained rock flour and coarse sediment, fluid overpressuring, and the steep nature of fjords make for a highly susceptible environment for failure during earthquakes, which are common in isostatically rebounding or tectonically uplifting regions where glaciers are likely.

The very nature of a fjord increases the landslide hazard. Fjords are an attractive place for human settlement as they offer good deepwater ports, flat land, and fresh water supply in mountainous regions close to the sea. Despite the fact that the Port of Valdez, Alaska (located at the head of a fjord) was destroyed following a submarine landslide that retrogressed onto land following an earthquake, the Trans-Alaska pipeline (which is responsible for bringing all of Alaska's North Slope crude oil to market) ends there. Were a similar event to have happened twenty years later, an unprecedented ecological disaster may well have occurred. To add to the hazard, the funnel-shaped geometry of fjords helps to focus tsunami waves. In an extreme example, a subaerial landslide at the head of a fjord in Lituya Bay, Alaska in 1958 triggered a 525 m high wave witnessed only by a few lucky boaters.

CANYONS

Submarine canyons are the primary transport avenues for sediment moving from land to the deep water. As such, these dynamic environments tend to have significant landslides. Canyon heads sequester the sediment that moves in littoral currents on the shelf. Following the 1989 Loma Prieta earthquake in central California, a failure occurred in the head of Monterey Canyon. These canyon-head failures likely fluidize (see FLUIDIZATION) into erosive turbidity currents that eroded the canyon floor, and are deposited on flat seafloor as turbidites. Successive incision events can eventually lead to large, slope-clearing landslides on the canyon walls (Densmore *et al.* 1997).

Regularly spaced canyons that occur on continental slopes but do not breach the shelf break, and are isolated from downslope erosive flows, are termed 'headless' (Orange and Breen 1992; see also GROUND WATER). Elevated seepage forces generated by the topography focusing groundwater flow to the seafloor can assist in failure, especially during transient events such as earthquakes (McAdoo *et al.* 1997). These failures are most likely small, but are a key factor in shaping the seafloor landscape.

CONTINENTAL SLOPE

Submarine landslides occur on continental slopes regardless of tectonic activity, latitude or sedimentation histories, however the temporal frequency of occurrence may vary substantially. The continental slopes are the regions in the ocean where the steepest sedimented slopes tend to exist, therefore it is a likely locus for landslide activity. The slope gradients, however, rarely approach the *angle of repose*, except in *canyons*, where *overconsolidated* (well-*lithified*) material is exposed at the seafloor, therefore either an increase in pore pressure or a significant transient is required for failure.

Many consider earthquakes the de facto trigger for slides on continental margins, even on passive margins where seismic events are less common. Curiously, one of the better concrete examples of an earthquake-triggered landslide occurred after the magnitude 7.2 Grand Banks earthquake in 1929 offshore from Canada's passive North Atlantic margin, and there have been many large earthquakes in regions with little evidence of substantial landsliding.

Landslides on open continental slopes may be associated with changes in sea level. Gas hydrate dissociation occurs during times of falling sea level, and may provide enough free gas to reduce the strength (increasing p_f) of the overburden, causing failure. Sedimentation rates are likely to be higher on the continental slope when sea level is low. Rivers bypass the sediment sinks in estuaries (see ESTUARY)and on the CONTINENTAL SHELF, and deposit material directly on the continental slope and in deep water by way of canyons. Rapid sedimentation not only increases pore pressures within the sediment, but provides a source of material to fail. Layers of finer grained mechanically weaker material deposited during highstands may provide a sliding surface following low stand sedimentation events. Continental slopes may well enter an equilibrium (see EQUILIBRIUM SLOPE) condition not long after a lowstand has provided material on the slope for which to fail.

VOLCANIC ISLANDS

Some of the largest and potentially most devastating submarine landslides occur on the flanks of volcanic islands and the mid-ocean ridges. Steep slopes are made of often highly fractured, hydrothermally altered and thermally stressed material. Earthquakes, rapid surface changes and rapidly expanding gases come with movement of subsurface magma. A block from prehistoric Nuuanu slide off Oahu, Hawaii is so large (30 km long, 17 km wide and 1.8 km thick) and far from its source (almost 100 km) it was initially identified as a seamount. A slide such as this may have been responsible for a tsunami that deposited material hundreds of metres high on the Hawaiian island of Lanai. Another slide on the Canary Islands in the Atlantic Ocean covered an area of 2,600 km^2 and 150 km^3 of material. It has been speculated that the tsunami from this slide was responsible for depositing a car-sized boulder some 20 m above sea level on Eleuthera Island, Bahamas on the other side of the ocean.

Hazard

Most submarine landslides are small, and are not likely to be noticed. The *hazard* associated with offshore slides is of primary concern to structures such as oil drilling rigs and production facilities, seafloor pipelines and cables. In 1969, Hurricane Camille in the Gulf of Mexico caused a drilling platform offshore from Louisiana to be buried in mud and moved 30 m downslope. Some slides begin in the offshore and can retrogress headward towards land. Following the 1964 Alaska 'Good Friday' earthquake, the Port of Valdez waterfront slumped into the ocean, killing thirty people.

TSUNAMI

Another significant *hazard* that may result from submarine landslides is TSUNAMI. There are two features that are unique to submarine landslide-generated tsunami. First, if the landslide was triggered by an earthquake, the tsunami generated tends to be larger than one might expect for the earthquake alone. Second, whereas earthquake-generated tsunami tend to affect the very large areas, and even cause significant damage throughout the ocean basin, landslide-generated tsunami tend to have more localized affects. On 17 July 1998 in Papua New Guinea, a tsunami with waves up to 15 m high followed a magnitude 7.0 earthquake, killing over 3,000 people. This was an unusually large tsunami given the size of the earthquake, and the region of devastation was limited to a small (~40 km) stretch of coast. Detailed multibeam and seismic surveys of the offshore revealed a large *amphitheatre* where a landslide likely triggered by the earthquake was initiated. This landslide probably contributed significantly to the amplitude of the tsunami.

The magnitude 7.2 earthquake offshore from Canada in 1929, triggered a large submarine landslide. This landslide quickly transformed into turbidity currents that travelled over 700 km into the abyssal plains, severing trans-Atlantic cables. Based on the timing of the cable breaks, the turbidity current velocity was calculated to be 55 km hr^{-1}. The seafloor displacement from the landslide caused a tsunami that hit Newfoundland's Burin Peninsula, killing twenty-eight people. Canada's worst earthquake-related disaster occurred on a passive margin because of a tsunami set off by a submarine landslide.

References

Bouriak, S., Vaneste, M. and Saoutkine, A. (2000) Inferred gas hydrates and clay diapers near the Storegga slide on the southern edge of the Voring Plateau, offshore Norway, *Marine Geology* 163(1–4), 125–148.

Densmore, Alexander L., Anderson, R.S., McAdoo, B.G. and Ellis, M.A. (1997) Hillslope evolution by bedrock landslides, *Science* 275(5,298), 369–372.

Hampton, M., Lee, H. and Locat, J. (1996) Submarine landslides, *Reviews of Geophysics* 34, 33–59.

Kayen, R. and Lee, H. (1991) Pleistocene slope instability of gas hydrate-laden sediment on the Beaufort Sea margin, *Marine Geotechnology* 10, 125–141.

Lee, H. and Edwards, B. (1986) Regional method to assess offshore slope stability, *Journal of Geotechnical Engineering, ASCE* 112, 489–509.

McAdoo, B., Orange, D., Screaton, E., Lee, H. and Kayen, R. (1997) Slope basins, headless canyons, and submarine palaeoseismology of the Cascadia accretionary complex, *Basin Research* 9, 313–324.

McAdoo, B., Pratson, L. and Orange, D. (2000) Submarine landslide geomorphology, US continental slope, *Marine Geology* 169, 103–136.

Mohrig, D., Whipple, K., Hondzo, M., Ellis, C. and Parker, G. (1998) Hydroplaning of subaqueous debris flows, *Geological Society of America Bulletin* 110(3), 387–394.

Moore, J. and Moore, G. (1984) Deposit from a giant wave on the island of Lanai, *Science* 226, 1,312–1,315.

Morgenstern, N.R. (1967) Submarine slumping and the initiation of turbidity currents, in A.F. Richards (ed.) *Marine Geotechnique*, 189–210, Urbana: University of Illinois Press.

Norem, H., Locat, J. and Schieldrop, B. (1990) An approach to the physics and the modeling of submarine landslides, *Marine Geotechnology* 9, 93–111.

Orange, D. and Breen, N. (1992) The effects of fluid escape on accretionary wedges II: seepage force, slope failure, headless submarine canyons and vents, *Journal of Geophysical Research* 97, 9,277–9,295.

Seed, H. and Rahman, M. (1978) Wave-induced pore pressure in relation to ocean floor stability of cohesionless soils, *Marine Geotechnology* 3, 123–150.

BRIAN G. McADOO

SUBMARINE VALLEY

Submarine valleys (canyons), sometimes tightly linked to fluvial systems, represent one of the main components that contribute to modelling the sea-bottom. They are one of the main ways for sediments to move from coastal zones to the deep oceanic basins, strongly cutting long bands of submarine scarp. Submarine valleys, localized on the main continental margins, show different features and evolutionary patterns according to whether they have a subaerial or submarine origin (De Pippo *et al.* 1999).

The characters and forms of some valleys confirm the hypothesis of their prevalently erosive origin as continental valleys. This hypothesis, even if it is acceptable for important palaeovalleys now located in shallow waters and filled with recent sediments, can hardly explain, except for particular cases, the characters and the evolution of active canyons. In fact, the bottom of most submarine valleys is a thousand metres deep, below the lowest depth (−200 m) reached by fluvial erosion during the maximum lowstand of sea level in the last glacial.

It is possible to distinguish the origin of submarine valleys on the base of certain main factors, such as eustatism and tectonics, that control their genesis and evolution. In particular the presence of valleys with a subaerial origin is more frequent in the less steep and deep areas that have experienced important sea-level changes, with the partial emersion of the continental shelf. Deep valleys, instead, are cut on steep continental margins where a big sedimentary supply activates big turbiditic flows that cut the trace of the primordial canyon.

In elevated depth areas, where the contrast between the sea waters and the waters filled with solutes and suspended sediments ceases, the mixture of water and sediment begins to slide offshore, also favoured by the shelf slope. The high sediment supply allows the erosion of the sediments previously deposited in the channels during the periods of insufficient solid supply by moving the sediments in the delta, from which turbiditic flows can originate following tectonic and/or seismic events. These flows are able to strongly erode the substratum, producing a deep incision on the shelf, where a valley could be already present. In shallow waters, instead, where the continental shelf temporarily emerges as a consequence of eustatic sea-level changes, the river directly affects the substratum, originating a submarine valley.

The tectonic characteristics of the area in which a canyon is present play a fundamental role in its genesis and evolution through time. At the same time the ampleness and the inclination of the shelf and continental slope, with the sedimentary mechanisms that regulate the transfer of the sediments from the shelf towards deeper basins, influence the evolutionary pattern of submarine valleys. In fact in basins fed by important fluvial systems, the sedimentary circulation is much more active than in areas where the only supply source is represented by gravitational processes. These processes are influenced, as well, by the eustatic sea-level changes, becoming more frequent in the lowstand periods than in the highstand ones.

The different morphologies observed in the canyons are tightly linked to different genesis and sedimentary processes. The cross section is one of the main morphological factors to be considered in the analysis of a canyon: the shape gives decisive information to define canyon evolution. A submarine valley with a V-shaped cross section generally points to dissection because of very speedy flows, while a U-shaped section points to the presence of slow and sporadic flows or to a relatively recent formation of the valley. The prevalence of depositional over erosional processes is another factor that contributes to canyon evolution by favouring the formation of a U-shaped, or at least, of a flared cross section. This causes the filling of the submarine valley bottom, with the consequent evolution of the canyon from a V towards a U-shape section. On the other hand, it is more frequent that the canyon presents steep walls and therefore a V-shaped section when erosion prevails over deposition. Slow phenomena, such as subsidence, can contribute to the evolution of the submarine valley cross section. In fact when canyons lie on slightly subsident basins they show a flared shape, while in more tectonically active areas, valleys frequently show more V-shaped sections.

The erosive processes in a canyon are strongly influenced by lithology. Canyons cut on very resistant bottom rocks, will tend in time to fill rather than become deeper. That happens in submarine valleys located along forearc systems, where the valley bottom shows a high resistance to erosion with a consequent widening of the valley section.

Another important aspect to be taken into account for the analysis of a canyon is its longitudinal profile. It can generally be pseudo-rectilinear to meandering with a high sinuosity index. In the areas that are not excessively steep that have been affected by tectonic events, but which are now stable, rectilinear canyons prevalently develop. The meandering ones, instead, generally develop in tectonically active areas and with more elevated gradients. The meandering of a canyon is function of its own form and longitudinal development. In fact the bending ray of a canyon and the meander wavelength increase both with width and inclination of the longitudinal profile. The presence of deep meandering channels points out the existence of transport processes characterized by frequent, continuous and slow turbiditic flows tightly connected to the inclination of the valley. In particular the local increment of canyon slope gradient is compensated by the increment of the profile sinuosity to maintain constant the solid supply.

It is possible to establish a relationship between the increment of slope gradient and the increment of the canyon sinuosity by the analysis of morphological parameters. Therefore the rectilinear or meandering profile of a submarine valley is mainly correlated both to the slope gradient value and to transport processes, and also to the nature of the material that flows inside.

The very rectilinear profile of a canyon can indicate that it has been cut on an important tectonic feature. Nevertheless valleys with a good rectilinear layout can change after tectonic events, because of continental shelf gradient variations, or also after eustatic sea-level changes.

In particular slope gradient variation owing to tectonic, tilting and subsidence processes, can cause the valley to evolve from a rectilinear towards a meandering profile. The evolution of a single submarine valley can also be strongly influenced both by the convergence and the feeding of the numerous canyons existing on the continental slope, causing consequent valley deepening and widening.

Reference

De Pippo, T., Hardi, M. and Pennetta, M. (1999) Main observations on genesis and morphological evolution of submarine valleys, *Zeitschrift für Geomorphologie* 43, 91–111.

SEE ALSO: canyon; rejuvenation; sea level

TOMMASO DE PIPPO

SUBMERGED FOREST

Submerged forests are former land surfaces on which *in situ* rooted tree stumps and associated organic deposits are found in the intertidal zone and on the continental shelves. They have been described by Geikie (1885), Reid (1913), Godwin (1943), Heyworth (1978) and Tooley (1979) in north-west Europe and by Krishnan (1982) and Mascarenhas (1997) in India.

Submerged forests owe their existence to a variety of causes: sea-level rise, coastal erosion, SUBSIDENCE and land uplift. The submerged forest at Rossall Beach on the Lancashire coast, United Kingdom, is the basal organic deposit of a kettle hole, the ramparts of which have been removed by erosion. Whereas at Hartlepool in north-east England (Tooley 1979) the submerged forest, which contains a Neolithic age human skeleton, is associated with estuarine clays. At Formby on the Lancashire coast the submerged forest formed in palaeo-dune slacks (Tooley 1979) and associated with them are human and animal footprints of Neolithic to Bronze Age (Huddart *et al.* 1999).

References

Geikie, A. (1885) *Textbook of Geology*, London: Macmillan.
Godwin, H. (1943) Coastal peat beds of the British Isles and North Sea, *Journal of Ecology* 31, 199–237.
Heyworth, A. (1978) Submerged forests, in J. Fletcher (ed.) *Dendrochronology in Northern Europe*, British Archaeological Reports S-51, 279–288.
Huddart, D., Roberts, G. and Gonzalez, S. (1999) Holocene human and animal footprints, Formby Point, NW England, *Quaternary International* 55, 29–41.
Krishnan, M.S. (1982) *Geology of India and Burma*, New Delhi: CBS.
Mascarenhas, A. (1997) Significance of peat on the western continental shelf of India, *Journal of the Geological Society of India* 49, 145–152.
Reid, C. (1913) *Submerged Forests*, Cambridge: Cambridge University Press.
Tooley, M.J (1979) Sea-level changes during the Flandrian stage, *Proceedings of the 1978 International Symposium on Coastal Evolution in the Quaternary*, Sao Paulo, Brazil, 502–533.

MICHAEL TOOLEY

SUBSIDENCE

Ground subsidence occurs only in specific environments, where any of three distinctive processes can occur. Compaction of porous and deformable clay, peat or silt causes surface subsidence as the soils restructure with declining pore space, normally accompanied by abstraction or expulsion of interstitial ground water. Collapse or deformation of the ground into natural caves occurs mainly in limestone, gypsum and basalt, and comparable dissolution occurs on salt. Large-scale processes include crustal sag, delta compaction, earthquake movements and volcanic deflation. Mining subsidence and the collapse of old mines are entirely artificial processes that are not a part of geomorphology.

Clays are deformable because their water is loosely bonded to the chemically complex clay mineral particles. They may therefore compact (and cause subsidence) when the water is squeezed out under load or in response to water abstraction. Clay subsidence may be regional or localized, but this ranks as the most widespread and most destructive subsidence process. Entire cities have subsided, with Venice (Plate 134), Mexico City, Tokyo, Shanghai and Bangkok among the better known examples.

The natural process of clay compaction causes water to be squeezed out during slow consolidation. Artificial removal of the water, by abstraction pumping, induces more rapid compaction. The subsidence hazard lies in alluvial sequences with alternating beds of poorly consolidated sand and clay beneath large cities. Convenient water supplies are pumped from the incompressible

Plate 134 Boardwalks across the Piazza San Marco in Venice, which has now subsided below the level of most winter high tides

sand aquifers, and the overall decline in pore-water pressures causes the adjacent clays to compact. The amount of subsidence is proportional to the groundwater head decline (Poland 1984), but greater subsidence occurs on younger clays that are less consolidated by self-weight, and also on those with higher contents of unstable smectite. This clay mineral forms primarily by weathering of volcanic rocks in wet tropical environments, and is a factor behind the major subsidence of Mexico City.

The role of pressure decline within overpumped aquifers is now recognized worldwide, and is clearly seen in the subsidence record of Venice. Long-term subsidence of the entire delta region combines with rising sea levels to create a continuing problem at Venice, but subsidence was greatly accelerated when groundwater abstraction caused clay compaction in response to head decline (Figure 163). Pumping controls have allowed head recovery, but 90 per cent of the induced compaction is irreversible, and the unabated natural subsidence continues to cause increasingly frequent flooding of the city.

Groundwater withdrawals can be matched by extraction of petroleum, natural gas and steam as causes of major ground subsidence, where the loss of hydraulic support may also affect rocks other than clay. Oilfield subsidence includes the classic case at Wilmington, USA, which caused the Long Beach harbour area of Los Angeles to subside by nearly 9 m.

Shrinkage due to water loss causes subsidence in any clay at shallow depths. Climatic changes

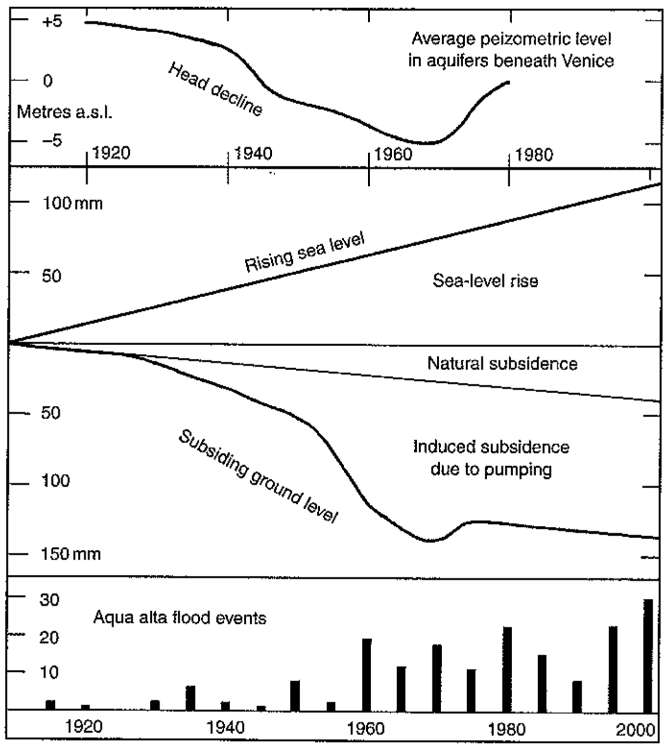


Figure 163 Causes and effects of the subsidence of Venice over the last 100 years. The bar graph shows, for each 5-year period, the numbers of flood events when high tides reach more than 600 mm above the level that initiates flooding in the lowest part of Piazza San Marco

affect large areas. Britain's dry summers since 1976 have produced a wave of subsidence damage to older houses on foundations so shallow that clay beneath them suffered first-time shrinkage in the new regime of drier climates. Subsidence is greater when trees extract excessive amounts of soil water during times of drought.

Subsidence on organic peat soils is induced by loading and/or drainage, and is similar to clay except that the compressibility of peat is far greater. Drainage causes immediate subsidence of peat, by about half the head decline, as recorded by the Holme Post in the English Fens (Hutchinson 1980). Through multiple phases of land drainage, the Post has emerged from the subsiding peat because it is founded on the stable clay beneath. After the first major movements, subsidence was about a quarter of head decline in each subsequent phase of drainage. Following the initial drainage and compaction, subsidence on peat continues due to wastage. This is loss to the atmosphere of the peat left above the water table and thereby exposed to microbial oxidation, causing annual surface lowering of 5–100 mm. In the drained English Fens, rivers have to be maintained between high banks across the subsidence bowls, land drainage water has to be pumped up into the elevated rivers, and buildings on piles progressively rise from the peat.

After clay, limestone KARST provides the world's most widespread subsidence environment, wherever it is at outcrop or beneath a soil cover. Caves constitute a subsidence hazard by threatening total ground collapse over small areas. The roof of a cave develops a natural arch in compression, which is stable when it is thinned down to about a quarter of its span (by surface lowering and roof stoping). Cave roofs need to be about double this thickness to take loads imposed by engineering, and variable degrees of fracturing and fissuring make cave roof stability difficult to assess. Also, locations of unseen caves are generally unpredictable, so cave collapse is a random and potentially destructive subsidence hazard. Most of the world's natural caves are in limestone, while gypsum and basalt are the other significantly cavernous rocks. However, collapses of rock over caves are very rare events; almost none of the world's limestone gorges originated as a collapsed cave.

A more widespread hazard is the development of subsidence dolines in soil covers that are washed into caves or fissures in underlying limestone. There are various types of DOLINES (alternatively known as sinkholes), each distinguished by its processes of erosion and collapse. Suffosion and dropout dolines (known collectively as subsidence dolines) form entirely within the soil profile (Plate 135); infiltrating rainfall washes soil down into pre-existing rockhead fissures at rates which can be significant to engineered structures. Slow slumping of non-cohesive sandy soils produces suffosion dolines that may damage structures but are not life-threatening. In a cohesive clay soil, cavitation initiates over a rockhead fissure, and grows slowly beneath an arched soil roof. It propagates upward until the surface collapses instantly and without warning; such a 'dropout' can be a major subsidence hazard in soil-covered karst, and individual failures can be up to 100 m across.

Subsidence sinkholes are the most frequent subsidence geohazard in karst terrains, especially the rapidly formed dropouts. Sites of new failures are impossible to predict, except that buried limestone boundaries with allogenic water input are recognizable subsidence zones, notably where soil cover is 1–15 m thick (though ground collapses have occurred where the cavernous limestone lies beneath 100 m of soil cover). Dropouts are caused by water flow, and therefore occur most frequently during rainstorms, and/or where drainage paths have been modified or water tables have declined due to over-abstraction. Most doline collapses are induced by engineering activity, and are therefore avoidable (Newton 1987).

Plate 135 A dropout doline at Ripon, Yorkshire. Only the soil profile is exposed after weak cover rocks and soil failed into cavernous gypsum beneath

Rock salt is rapidly soluble in natural waters. Total removal of salt beds can develop within a few years, orders of magnitude faster than mineral losses in limestone, and ground subsidence occurs at significant rates where salt is being removed by circulating ground water. In lowland sites, including the Cheshire saltfield of England, dissolution at the rockhead leaves insoluble material in unstable residual breccias beneath the drift cover. Rates of dissolution and subsidence are a function of groundwater flow patterns, and overall natural subsidence is generally $>0.1\,\mathrm{mm\,yr^{-1}}$; during the Pleistocene, this created subsidence hollows that are now occupied by the mere lakes. All salt subsidence is vastly accelerated by any brine-pumping operations that draw in new supplies of chemically aggressive fresh water to replace the abstracted brine. Traditional wild brining targets the 'brine streams' that are zones of enhanced groundwater flow just above rockhead, and thereby causes the linear subsidences above them to deepen by $100\,\mathrm{mm\,yr^{-1}}$ or more, ultimately creating new ribbon lakes known as 'flashes'.

It is significant that all the widespread subsidence processes on clay, limestone, gypsum and salt are induced or accelerated by human activities. Subsidence over active and abandoned mines is entirely due to humans. The implication therefore stands that subsidence is a process and a geohazard that can largely be controlled. The same applies to other, less common types of subsidence.

Collapsing soils are sediments prone to internal structural collapse when water is added to them. Weak clay bonds between the grains of loosely packed, silt sediments are broken by the introduction of water; the soil then densifies by repacking under self-load or imposed load, in the process known as hydrocompaction. Potentially collapsible soils are wind-deposited LOESS and some alluvial silts that were rapidly deposited and then desiccated in large basinal fans. Hydrocompaction can promote ground subsidence of up to 5 m over wide areas in semi-arid regions where the soils have not been previously wetted.

Clays and silts are generally weak when wet and saturated, but become solid when frozen in PERMAFROST. Subsidence is then inevitable where ice-rich soils are thawed by stripping vegetation or placing warm buildings directly on the ground, though sands and gravels are largely thaw-stable.

The only subsidence that is totally natural, and therefore not controllable, is that originating with deep-seated processes. True tectonic subsidence occurs over boundary zones of subduction and also where plates are necked and thinned in zones of tension; the London area is subsiding by more than $2\,\mathrm{mm\,yr^{-1}}$ due to thinning and sinking of the North Sea Basin. Such rates of subsidence may be critical at coastal sites, especially when combined with the current rise in world sea levels (at 1–2 $\mathrm{mm\,yr^{-1}}$), largely due to global warming that has been continuous for about the last 500 years.

Major deltaic basins are sites of very slow subsidence due to a combination of factors. Within the sediment piles, soft clays are consolidated into mudstones with reduced porosities, and this compaction causes subsidence. At the same time crustal sag of the overburdened basin floor provides a second component of subsidence. Venice lies towards the margin zone of the Po Valley deltaic basin, where long-term mean subsidence is $0.4\,\mathrm{mm\,yr^{-1}}$ due to both compaction and sag; this is the natural, uncontrollable component of Venice's continued subsidence (Figure 163).

Accelerated tectonic subsidence is accompanied by earthquakes when crustal deformation is transferred to fault displacement. Large areas in Alaska subsided by up to 2.5 m during the 1964 earthquake, causing permanent drowning of coastal forests. Subsidence is more widespread as a secondary effect of earthquakes, due to liquefaction of unconsolidated sand soils. During the period of earthquake vibration, sands may temporarily behave as a liquid, so that structures subside into them. The vibrations also cause densification of loose sands by improved grain packing, and this causes permanent subsidence of the ground surface.

Entire volcanoes, and their immediate surrounds, can deflate due to migration or retreat of magma from beneath them. Parts of the Naples region subsided by 10 m due to 1,200 years of deflation of the Campi Flegri volcanic centre. This is one case where subsidence is welcome, as the reverse uplift is due to volcanic inflation that normally precedes an even more destructive eruption.

References

Hutchinson, J.N. (1980) The record of peat wastage in the East Anglia fenlands at Holme Post, 1846–1978 AD, *Journal of Ecology* 68, 229–249.

Newton, J.G. (1987) Development of sinkholes resulting from man's activities in the eastern United States, *US Geological Survey Circular* 968, 1–54.

Poland, J.F. (ed.) (1984) *Guidebook to Studies of Land Subsidence due to Groundwater Withdrawal*, Unesco Studies and Reports in Hydrology 40.

Further reading

Carbognin, L. and Gatto, P. (1986) An overview of the subsidence of Venice, *International Association Hydrological Sciences Publication* 151, 321–328.

Cooper, A.H. and Waltham, A.C. (1999) Subsidence caused by gypsum dissolution at Ripon, North Yorkshire, *Quarterly Journal of Engineering Geology* 32, 305–310.

Culshaw, M.G. and Waltham, A.C. (1987) Natural and artificial cavities as ground engineering hazards, *Quarterly Journal of Engineering Geology* 20, 139–150.

Holzer, T.L. (1991) Nontectonic subsidence, *Geological Society America Centennial Special Volume* 3, 219–232.

Prokopovich, N.P. (1986) Origin and treatment of hydrocompaction on the San Joaquin Valley, USA, *International Association Hydrological Sciences Publication* 151, 537–546.

Stephens, J.C., Allen, L.H. and Chen, E. (1984) Organic soil subsidence, *Geological Society America Reviews Engineering Geology* 6, 107–122.

Waltham, A.C. (1989) *Ground Subsidence*, Glasgow: Blackie.

TONY WALTHAM

SUFFOSION

An erosional process, occurring in areas where well developed KARST is overlain by unconsolidated superficial materials (usually loess or till). Suffosion is associated with PIPES AND PIPING, and is also known as ravelling, describing both catastrophic and gradual collapse of the superficial material into the bedrock cavity. The sediments slump down into widened joints and cavities in the bedrock surface, producing an irregular land surface, often exhibiting dimpling and multiple suffusion dolines (also termed shakeholes). The unconsolidated sediments can be susceptible to collapse compression upon saturation, with infiltrating water beneath the regolith creating subsoil KARREN and widened joints connected with deeper cavities. Fine materials are often eroded internally by a combination of solution and downwashing (mechanical suffosion), suggesting that these processes are related to karst underwatering. Layers of more compacted material may arch over the collapsing soil, and these are often attributed to lowering of the water table. Suffosion dolines can form within seconds and are the most widespread problem encountered in karst terrains in the construction industry, save pollution of aquifers. The depressions thereby formed typically range from 1 m diameter and 0.5 m deep, up to 100–200 m diameter and 10–50 m deep (though these are less frequent) (Ford and Williams 1989: 525).

Reference

Ford, D. and Williams, P. (1989) *Karst Geomorphology and Hydrology*, London: Unwin Hyman.

SEE ALSO: doline

STEVE WARD

SULA

Rocky zones along stream courses, especially in the tropics, where channels divide into anastomosing channels associated with cataracts (Zonneveld 1972). They have been explained in terms of rivers crossing an irregular weathering front, having limited amounts of abrasive material to enable incision, the induration of outcrops with ferromanganese compounds, and possible inheritance of braided conditions from Quaternary dry periods. The view that tropical rivers lack abrasive tools because of the speed and thoroughness of weathering reducing the amount and size of sediment load has been advanced by many climatic geomorphologists (e.g. Büdel 1982), but reliable data on this are sparse.

References

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

Zonneveld, J.I.S. (1972) Sulas and sula complexes, *Göttinger Geographische Abhandlungen* 60, 93–101.

A.S. GOUDIE

SULPHATION

A reaction between sulphur dioxide from the atmosphere and building materials (including stone) to form gypsum (calcium sulphate). It is often regarded as a feature of polluted urban atmospheres. Sulphur dioxide becomes oxidized (sometimes in the presence of catalysts such as soot or metal-rich particles) on moist surfaces to form sulphuric acid. The sulphuric acid then reacts with the stone in the following way:

$$CaCO_3 + H_2SO_4 + 2H_2O \rightarrow CaSO_4{\cdot}2H_2O + CO_2 + H_2O$$

According to Cooke and Gibbs (1993), where little moisture is present two alternative series of reactions may occur. Hydrated calcium sulphate may form first from the reaction of calcium carbonate and sulphur dioxide, as represented in the following reaction:

$$CaCO_3 + SO_2 + 2H_2O \rightarrow CaSO_4{\cdot}2H_2O + CO_2$$

Subsequently, this hydrated calcium sulphate may become oxidized to form gypsum in the presence of catalysts. Alternatively, sulphur dioxide may react with bicarbonate solutions formed by reaction with rainfall containing dissolved carbon dioxide as follows:

$$Ca(HCO_3)_2 + SO_2 \rightarrow CaSO_3 + 2CO_2 + 2H$$

The deposition of sulphates on material surfaces, often in the form of a normally dark gypsum crust, can be seen as both a consequence and a cause of weathering. This is especially true of limestones, where such a crust may be harder than the material on which it has developed. Other properties may also be different and may cause an acceleration in the speed of stone degradation. Amoroso and Fassina (1983: 264) identify three mechanisms that may account for this. First, a variation in volume occurs because gypsum has a greater volume than the quantity of calcite it replaces. One volume of calcium carbonate forms over two volumes of hydrated calcium sulphate. This causes expansive stresses in pores and cracks. Second, calcite and gypsum have different thermal expansion characteristics. The linear coefficient of the thermal expansion of gypsum is about five times that of calcite. This difference may be further increased when blackened crusts develop because they tend to absorb a larger amount of radiation than white surfaces. Third, the development of a crust will reduce the permeability of the material, which will in turn increase the water retention beneath.

Sulphation can lead to blister development and lamination, possibly through a combination of the factors just described.

Gypsum crusts in urban environments can also develop on non-carbonate rocks such as sandstone, either as a result of sulphation of calcium derived from an external source (e.g. the mortar surrounding the stone), or because of the accumulative role of lichens, algae, bacteria, etc., or because of chemical reactions with certain minerals within the rock (McKinley *et al.* 2001).

References

Amoroso, C.G. and Fassina, V. (1983) *Stone Decay and Conservation*, Materials Science Monographs 11, Amsterdam: Elsevier.

Cooke, R.U. and Gibbs, G.B. (1993) *Crumbling Heritage? Studies of Stone Weathering in Polluted Atmospheres*, London: National Power and Powergen.

McKinley, J.M., Curran, J.M. and Turkington, A.V. (2001) Gypsum formation in non-calcareous building sandstone: a case study of Scrabo Sandstone, *Earth Surface Processes and Landforms* 26, 869–875.

A.S. GOUDIE

SUPRAGLACIAL

Supraglacial refers to the surface zone of glaciers and ice sheets where there is distinctive drainage, sediment sources, transport and deposition and a set of landforms characteristic of the supraglacial association and landsystem (Paul 1983). Many of the landforms produced are ephemeral and do not survive deglaciation but some do, particularly where associated with stagnant ice.

Supraglacial debris enters glacial transport from a number of sources such as mass movements from nearby mountains, tephras and from subglacial sediment carried to the surface. Minor inputs include meteorites, pollutants from anthropogenic sources and sea spray salts. Usually in high mountain glaciers where the surface ice is adjacent to valley slopes, or where isolated peaks (nunataks) protrude through the ice sheet, mass movements from the slopes are the dominant surface sediment supplier. These processes include rockfall, slides, snow and ice avalanching, debris flows, slushflows and streamflows. The supplied supraglacial debris is incorporated into the glacier by snow and ice burial in the accumulation area and by falling into crevasses and MOULINs (cylindrical vertical shafts). This is then transported by high-level, passive debris transport and suffers little modification, retaining its primary, parent material characteristics. These are angular to very angular particle shape, with common elongate to slabby particles, and coarse-grain size, with little fines.

Surface snowmelt allows downward water percolation through the snowpack and if the melting exceeds refreezing, water accumulates as slush swamps. Usually water drains down a gradient forming rills and eventually a dendritic drainage network. If discharge is high enough surface channels form, being well developed on the ablation area ice because of its low permeability. These channels are smooth-sided, offering minimal water flow resistance, hence high velocities and may follow structural weaknesses, like foliation. On warm ice the surface channels are usually short and water is diverted into the glacier via crevasses and moulins. On cold glaciers supraglacial channels commonly flow towards and alongside the margins. Supraglacial lakes can form during the early ablation season but in temperate glaciers they drain as the englacial drainage opens. In cold ice these lakes can persist and where there are large amounts of supraglacial debris, differential melting causes widespread formation of water-filled hollows. Supraglacial lakes, or ice-cauldrons, also form in depressions created by geothermal melting and subsidence.

Supraglacial debris is spatially variable and it has an important control on ablation dynamics. It acts as insulation and retards surface melting. Differential ablation causes dirt cone development (Drewry 1972) and glacier tables can form by boulders protecting a pedestal of ice from melting. Supraglacial lateral MORAINEs are ice-cored, debris accumulations at valley glacier margins, often formed by scree cone coalescence. Ablation reduction by thick debris cover means that these landforms stand above the adjacent relatively clean ice. Medial moraines are the supraglacial expression of a vertical medial debris septum which can extend to the glacier base but the surface debris is usually more concentrated and laterally extensive than that within the glacier because glacier ablation redistributes the debris. They have been classified into several types by Eyles and Rogerson (1978), including ice stream interaction medial moraines and ablation-dominated medial moraines. The former are created by lateral moraine confluence at the junction of two glaciers where the medial moraine marks the suture line between the glaciers. The second type form where ridges of englacial debris are revealed downglacier by ablation. They can form in two ways. A rockwall on a nunatak can supply debris to a small glacier area forming a linear debris plume downglacier from that point. This debris descends in the accumulation zone in the flow direction as it becomes buried. In the ablation zone it is revealed as a supraglacial medial moraine. Alternatively, folding of debris-rich, ice stratification may occur, particularly where ice flows into a restricted channel (Hambrey *et al.* 1999). The axes of these folds are usually parallel to the ice flow direction. When the debris-rich ice reaches the ablation zone, surface melting reveals the anticlinal crests of the longitudinal, debris-rich folds. If the folding is relatively open then a series of small medial moraines may define the axis of a fold.

Some glaciers have much supraglacial debris mantling the lower part of their ablation zone, particularly where mass wasting delivers large debris volumes in both the accumulation and ablation areas, as in high relief mountains, or where topography allows the movement of basal debris to the surface along shear planes. Uneven reworking and debris deposition in multiple events during ablation is responsible for a distinctive landform assemblage which can lead to topographic inversion on the glacier surface. The result is a complex sediment assemblage of faulted and folded fluvial, lacustrine and mass movement sediment (Benn and Evans 1998; Huddart 1999) which is finally deposited on the substrate. This can include supraglacial ESKERs which can be let down from supraglacial and englacial channels, ice-walled lake plains (Huddart 1983) and several KAME types. Supraglacial debris can also contribute to the GLACIMARINE environment as part of morainal banks and from iceberg rafting.

References

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, London: Arnold.

Drewry, D.J. (1972) A quantitative assessment of dirt-cone dynamics, *Journal of Glaciology* 11, 431–446.

Eyles, N. and Rogerson, R.J. (1978) A framework for the investigation of medial moraine formation: Austerdalsbreen, Norway and Berendon Glacier, British Columbia, *Journal of Glaciology* 20, 99–113.

Hambrey, M.J., Bennett, M.R., Dowdeswell, J.A., Glasser, N.F. and Huddart, D. (1999) Debris entrainment and transfer in polythermal valley glaciers, *Journal of Glaciology* 45, 69–86.

Huddart, D. (1983) Flow tills and ice-walled lacustrine sediments, the Petteril Valley, Cumbria, England, in E.B. Evenson, C. Schlüchter and J. Rabassa (eds) *Tills and Related Deposits*, 81–94, Rotterdam: Balkema.

——(1999) Supraglacial trough fills, southern Scotland: origins and implications for deglacial processes, *Glacial Geology and Geomorphology*, 1–16, http://boris.qub.ac.uk/ggg/papers/full/1999/rp041999/rp04.html

Paul, M.A. (1983) The supraglacial landsystem, in N. Eyles (ed.) *Glacial Geology*, 71–90, Oxford: Pergamon Press.

SEE ALSO: esker; glacial deposition; ice stagnation topography; kame; moraine; moulin

DAVID HUDDART

SURGING GLACIER

Surging glaciers are often instantly recognizable from their 'looped medial moraines' or teardrop-shaped loops of debris on the glacier surface. These loops record cyclic changes in velocity between a trunk GLACIER and its tributaries. Additionally, folding, thrust faulting and severe crevassing of ice at the snout is induced by compressive flow. A glacier is said to surge when it exhibits major fluctuations in velocity over timescales that range from a few years to several centuries, swinging between phases of rapid and slow flow. The phase of rapid motion is called the 'surge' or 'active' phase during which ice is transferred from the upper part of the glacier (the reservoir area) to the snout. This results in a dramatic advance of the snout and a concomitant thinning of the reservoir area. The period of slow flow between surges is termed the 'quiescent phase' and is characterized by the build-up of ice in the reservoir area and the mass stagnation of the snout. This gradually increases the surface gradient of the glacier until the next surge is initiated. Glacier flow velocities in the surge phase can be ten times faster than those of the quiescent phase. The surge and quiescent phases are usually of constant length for each individual glacier, resulting in a periodic cycle of surges. However, cycle length varies greatly between glaciers and glaciated regions. Glacier surge velocities also vary considerably, ranging from 50 m per day on the Variegated Glacier in Alaska to maximum speeds of only 16 m per day on Svalbard. The length of the quiescent phase also varies, maximum periods of time being 50–500 years for Svalbard glaciers compared to 20–40 years for most other regions.

Geographically, surging glaciers tend to cluster in Alaska, Yukon and British Columbia in North America, Svalbard and Iceland in the North Atlantic and the Pamirs in western Asia. Further examples have, however, been identified in the Canadian high arctic, Greenland, the Caucasus, Tien Shan and Karakoram mountains in Asia and in the Andes. Several glacier characteristics in these regions have been correlated with surging behaviour, such as glacier length and overall gradient. Bedrock types also appear to be linked to surging.

Glacier surges are not triggered by climatic fluctuations but instead result from changes in the internal dynamics of the glacier system. Variations in basal sliding appear to drive the changes in glacier flow velocity in a surging glacier, and this is driven by reorganizations of the subglacial drainage system. Based upon the few intense studies undertaken on surging glaciers, rapid sliding appears to be initiated by rising water pressures at the glacier bed. This is a function of the trapping of subglacial meltwater by conduit closure due to ice creep. Although this mechanism of fast flow initiation applies well to temperate glaciers it does not explain fully surges by subpolar glaciers. A study on Trapridge Glacier in the Yukon, which is frozen to its bed at the snout, suggests that deformable sediment beneath the warmer ice further up glacier plays a significant role in surging (Clarke *et al.* 1984). A wave-like bulge develops on the glacier surface at the boundary between cold and warm-based ice. This bulge moves down the glacier mostly by subsole deformation. This is initiated by changes in the subglacial water flow through the deformable substrate.

A linkage between surging behaviour and regional climate has been suggested by Budd (1975), who indicates that the concentration of surging glaciers in specific regions, and the variability of surge velocities and phases between regions, must reflect some climatic control. Specifically, in a continuously fast-flowing glacier the annual mass balance can be discharged by normal flow velocities. In contrast, in a surging glacier the mass throughput is too great to be discharged by slow flow alone but too small to sustain fast flow over long periods. Consequently, surging glaciers build up mass slowly until fast flow is triggered. This fast flow drains and exhausts the supply from the reservoir, thereby reinitiating slow flow. This means that the velocity of a surging glacier is constantly out of equilibrium with climate.

References

Budd, W.F. (1975) A first simple model of periodically self-surging glaciers, *Journal of Glaciology* 14, 3–21.
Clarke, G.K.C., Collins, S.G. and Thompson, D.E. (1984) Flow, thermal structure and subglacial conditions of a surge-type glacier, *Canadian Journal of Earth Sciences* 21, 232–240.

Further reading

Benn, D.I. and Evans, D.J.A. (1998) *Glaciers and Glaciation*, 169–175, London: Arnold.
Clarke, G.K.C., Schmok, J.P., Ommaney, C.S.L. and Collins, S.G. (1986) Characteristics of surge-type glaciers. *Journal of Geophysical Research* 91, 7,165–7,180.
Dowdeswell, J.A., Hamilton, G.S. and Hagen, J.O. (1991) The duration of the active phase of surge-type glaciers: contrasts between Svalbard and other regions, *Journal of Glaciology* 37, 388–400.
Fowler, A.C. (1987) A theory of glacier surges, *Journal of Geophysical Research* 92, 9,111–9,120.
Jiskoot, H., Murray, T. and Boyle, P. (2000) Controls on the distribution of surge-type glaciers in Svalbard, *Journal of Glaciology* 46, 412–422.
Kamb, B., Raymond, C.F., Harrison, W.D., Engelhardt, H., Echelmeyer, K.A., Humphrey, N., Brugman, M.M. and Pfeffer, T. (1985) Glacier surge mechanism: 1982–1983 surge of Variegated Glacier, Alaska, *Science* 227, 469–479.
Lawson, W., Sharp, M. and Hambrey, M.J. (1994) The structural geology of a surge-type glacier, *Journal of Structural Geology* 16, 1,447–1,462.
Meier, M.F. and Post, A.S. (1969) What are glacier surges? *Canadian Journal of Earth Sciences* 6, 807–819.
Murray, T. and Porter, P.R. (2001) Basal conditions beneath a soft-bedded polythermal surge-type glacier: Bakaninbreen, Svalbard, *Quaternary International* 86, 103–116.
Raymond, C.F. (1987) How do glaciers surge? A review, *Journal of Geophysical Research* 92, 9,121–9,134.
Raymond, C.F. and Harrison, W.D. (1988) Evolution of Variegated Glacier, Alaska, USA, prior to its surge, *Journal of Glaciology* 34, 154–169.
Sharp, M.J. (1988) Surging glaciers: behaviour and mechanisms, *Progress in Physical Geography* 12, 349–370.

DAVID J.A. EVANS

SUSPENDED LOAD

The suspended load of a river comprises mineral and organic matter that is dispersed through the flow by turbulence. Typically, the mineral load is dominant and consists of grains ranging in size from clay (<4 μm) up to sand grade (<2 mm). Often, suspended particles exist as 'flocs' or clumps of finer grade particles linked by weak chemical forces. The suspended load is quantified in terms of its concentration (mass of sediment per unit volume of water), discharge (sediment mass flux per unit time – also referred to as the 'load'), and particle-size distribution (proportions of the load in given size fractions).

The clay-silt fractions (often termed 'washload') are largely sourced from erosion processes outside the river channel; being more easily suspended, they are well mixed through the flow and travel long distances in suspension. Sand requires more intense turbulence to remain suspended, thus it tends to be concentrated near the stream bed and tends to be sourced from the stream-bed material. In cobble/boulder-bed rivers, the suspended load is also augmented by fine particles generated by abrasion of cobbles/boulders as they roll, hop and slide along the bed as part of the BEDLOAD. Downstream through a drainage basin, the suspended load increasingly dominates over the bedload.

Since the fine washload is easily suspended, its concentration generally depends only on the relative rates of supply of sediment and water to the channel (i.e. rather than being limited by the physical capacity of the stream to suspend it). Indeed, when the concentration of mud-rich washload becomes sufficiently large, the fluid properties change from those of a water flow to those of a hyperconcentrated or debris flow. While turbulence intensity does limit the suspended sand concentration in sand-bed streams, in gravel or rock-bed channels the primary limitation on suspended sand concentration may be sand availability. Thus in most rivers, the suspended load tends to be 'under-supplied' with respect to the river's potential capacity to transport it. Because of this, it cannot be determined by physics-based formula but must be measured.

Accurate suspended load measurement at a river section requires measurements of both sediment concentration and water velocity. One approach is to collect point samples of water and to make point velocity measurements at intervals over the flow depth, then plot profiles of concentration and velocity. Special 'point-samplers' are used for this purpose. These cable-suspended samplers, comprising a brass bomb and an internal sample bottle, have a solenoid-operated valve to control the water-sample capture and they are designed so that they sample isokinetically (i.e. they accept a sample at the ambient water velocity). A second, simpler approach is to use a 'depth-integrating' sampler. With this, the inlet nozzle is kept open while the sampler is traversed from the water surface to the bed and back again, so performing a mechanical integration of the concentration and velocity profile. Details about samplers and methods developed in North America can be found in Edwards and Glysson (1999).

A single suspended sediment measurement is time consuming, and where a continuous record is required it is usually obtained by collecting 'index' samples at one location and establishing a relationship between the index sample concentration and the cross-section mean concentration. Index samples may be collected manually, but in remote locations or in 'flashy' small basins they are more often collected by an automatic pumping-sampler or by measuring a surrogate property such as water turbidity. Turbidity sensors have been widely used to date (e.g. Gippel 1995). They may be either transmissivity (attenuance) or back-scattering (nephelometric) types. Since they do not sense sediment concentration directly, a further calibration relation must be established between the turbidity sensor and suspended sediment concentration. The turbidity signal depends both on sediment concentration and particle characteristics, notably particle size and shape. Light scattering is greatest from clay particles and least from sand grains, thus turbidity sensors are more sensitive to the washload. Suspended sediment concentrations at-a-site tend to increase with water discharge, and so a relatively simple way to estimate the suspended discharge over a period of time is with a SEDIMENT RATING CURVE that empirically predicts sediment concentration at a given water discharge.

Globally, average annual river suspended sediment discharges vary greatly. The primary factors causing this variation include basin area, topography, precipitation, lithology, tectonics, vegetation cover, land use, floodplain sequestration and sediment entrapment in reservoirs. On a unit area basis, the largest sediment discharges occur in tectonically active, high-rainfall, steeplands around the western Pacific basin (e.g. Taiwan and New Zealand, where sediment discharges can exceed $20{,}000\,t\,km^{-2}\,yr^{-1}$). According to Milliman and Syvitski (1992), the largest sediment discharges from single river basins to the oceans are from the Amazon River ($1.2 \times 10^9\,t\,yr^{-1}$, due to its vast area) and China's Huanghe River ($1.1 \times 10^9\,t\,yr^{-1}$, due to its large areas of landuse-affected erodible loess terrane). Milliman and Meade (1983) estimated the total worldwide delivery of suspended sediment to the oceans at $13.5 \times 10^9\,t\,yr^{-1}$. Milliman and Syvitski (1992) revised this figure and considered that it might have been at least $20 \times 10^9\,t\,yr^{-1}$ before the proliferation of dams during the twentieth century, which have intercepted a significant fraction of the sediment discharge. They considered that global suspended sediment discharges have more than doubled as a result of widespread deforestation and farming over recent millennia.

References

Edwards, T.K. and Glysson, G.D. (1999) Field methods for measurement of fluvial sediment, *US Geological Survey Techniques of Water-resources Investigations* Book 3, Chapter C2.

Gippel, C.J. (1995) Potential of turbidity monitoring for measuring the transport of suspended solids in streams, *Hydrological Processes* 9, 83–97.

Milliman, J.D. and Meade, R.H. (1983) World-wide delivery of river sediment to the oceans, *Journal of Geology* 91, 1–21.

Milliman, J.D. and Syvitski, J.P.M. (1992) Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountainous rivers, *Journal of Geology* 100, 525–544.

Further reading

Hicks, D.M. and Gomez, B. (2003) Sediment transport, in G.M. Kondolf and H. Piegay (eds) *Tools in Fluvial Geomorphology*, 425–461, San Francisco: Wiley.

D. MURRAY HICKS

SYNGENETIC KARST

This refers to karst that has developed concurrently with the diagenesis and consolidation of the host karst rock. It is most well known in calcareous dune limestones, rocks which are largely of biogenic origin, being comminuted shell fragments consolidated into calcareous AEOLIANITE, although there may be significant admixtures of quartz and other non-calcareous minerals. Such rocks are usually found within 40° north and south of the Equator and karst development in them is especially well known in the coastal regions of South and West Australia (White 2000), but is also found in other regions such as the eastern and southern Mediterranean and on Caribbean islands.

The dunes developed during glacio-eustatic oscillations during the Quaternary that exposed large amounts of shell material in coastal regions. Dunes were blown inland and accumulated in a series of ridges roughly parallel to the coast. The first stage in diagenesis was fixing by vegetation, the growth of which promoted soil development and production of biogenic carbon dioxide in the root zone. This encouraged dissolution of the dominantly aragonitic sands by infiltrating rain water. Saturation with respect to calcium carbonate is achieved close to the surface, under conditions that are open system with respect to carbon dioxide. Further percolation of this saturated solution down into the sands away from the root zone results in degassing of carbon dioxide and evaporation, with the result that the water becomes supersaturated and calcite is precipitated in interstitial pores in the dune, thus cementing it. Cementation produces a surface case-hardened crust over the dune that advances progressively downwards from the surface. Vertical solution pipes with case-hardened rims often perforate the dune crust. Cementation also occurs in the water saturated (phreatic) zone, and eventually the entire dune becomes calcreted.

During this process of dune cementation streams flow from inland areas towards the coast, but are blocked in their progress by dune ridge barriers that lie parallel to the coast. The streams pond on the upstream side and their waters permeate the porous sands to emerge on the seawards side. As the dune sands become indurated, stream flow is restricted to defined routes which eventually develop into blind valleys leading into caves many of which have flat water-table controlled roofs. The cave ceilings are also case hardened. Some cave ceilings collapse and give rise to large chambers that become profusely decorated with speleothems. Further collapse develops collapse dolines.

Reference

White, S. (2000) Syngenetic karst in coastal dune limestone: a review, in A.B. Klimchouk, D.C. Ford, A.N. Palmer and W. Dreybrodt (eds) *Speleogenesis: Evolution of Karst Aquifers*, 234–237, Huntsville, AL: National Speleological Society.

PAUL W. WILLIAMS

SYSTEMS IN GEOMORPHOLOGY

Students of the Earth's surface have been referring to 'systems' for nearly three centuries: but when Buffon (1749) did so, he was concerned with the theoretical structures (which we should now term cosmogonies) of the likes of Whiston, Burnet and Woodward; whereas Playfair (1802: 102) in his famous discussion of the 'system of vallies' was dealing with concrete entities. These two aspects were brought together in A.N. Strahler's enormously influential paper on the dynamic basis of geomorphology (1952). Here Strahler considers (pp. 934–935) that the 'fullest development' of geomorphology will occur 'only when the forms and processes are related in terms of dynamic systems and the transformation of mass and energy are considered as functions of time'. He goes on (p. 935): 'Many of the geomorphic processes operate in clearly defined systems that can be isolated for analysis.'

Strahler's view derived from early papers by the biologist, Ludwig von Bertalanffy, notably that published in *Science* in 1950. The move towards the use of systems as a theoretical device, unifying all sciences and extending to fields such as economics, was christened General Systems Theory (GST) by von Bertalanffy (1950). Both the theoretical and the practical aspects of GST were adopted by several of Strahler's students, most notably by the British geomorphologist R.J. Chorley (1960; Chorley and Kennedy 1971) and the American S.A. Schumm (1977). The two versions of 'system' can be seen side by side in the Twenty-third Binghamton Symposium (Phillips and Renwick 1992).

So what *are* systems? and what has been their dual role in geomorphology since 1952? According to Hall and Fagan, 'A system is a set of objects together with relationships between the objects and between their attributes' (1956: 18). They acknowledge that this is an imprecise definition, but say 'This difficulty arises from the concept we are trying to define; it simply is not amenable to complete and sharp description' (*idem*). Further, they acknowledge the existence of abstract or conceptual systems (pp. 19–20) as well as natural systems (pp. 23–24), which latter they consider as either open (exchanging 'materials, energies or information') with their surroundings; or closed (with no such imports and exports). Chorley and Kennedy (1971: 2) modified this distinction by relabelling Hall and Fagan's 'closed' systems as 'isolated' and identifying a new category of 'closed' system which exchanged energy (or, presumably, information) but not mass with its environment. In this sense, there must be very few true isolated systems studied by geomorphologists, but there may be several categories of effectively closed systems, including the planet Earth itself. Chorley and Kennedy went on (pp. 4–10) to identify four classes of system of relevance to physical geography: morphological, cascading, process-response and control. They discuss each in turn, devoting most attention to the process-response examples together with related considerations, including input and output, equilibrium, thresholds (see THRESHOLD, GEOMORPHIC), trends and simulation models. Finally, Chorley and Kennedy consider control systems, a topic taken up and elaborated by Bennett and Chorley (1978) in a book ranging far beyond geomorphology.

So how have these ideas – of both the conceptual and the natural systems – actually been employed by geomorphologists? Chorley (1960) followed Strahler's (1952) lead in emphasizing the distinction between historically focused geomorphology (explicitly, the ideas of the American colossus, W.M. Davis and, in particular, his concept of the geographical cycle, see CYCLE OF EROSION), which he termed 'closed system' thinking: and the 'open system' approach deriving from GST and the study of process. Chorley lists (1960: B8 and 9) seven advantages which will accrue from the adoption of the 'open system' view. They are:

1 Emphasis will be placed on the 'tendency to adjustment' between form and process.
2 Emphasis will be directed towards 'the essentially multivariate character' of geomorphic processes.
3 It will allow consideration to be given both to progressive changes and to stasis or abrupt changes (i.e. both concepts such as DYNAMIC EQUILIBRIUM and thresholds).
4 It will foster 'a less rigid view regarding the aims and methods of geomorphology' than (he contends) is held by those concerned with studying the historical development of landscapes and landforms.
5 Emphasis will be placed upon the whole landscape, rather than on 'the often minute elements ... having supposed evolutionary significance' (by this he clearly has in mind the 1950s' obsession with terrace and PENEPLAIN remnants).
6 It will encourage 'rigorous' work in areas where there are few or no traces of erosional history.
7 There will be major implications for geography as a whole, especially by directing attention 'to the increasingly hierarchical differentiation which often takes place with time'.

More than half a century on, it is not evident that these advantages have been realized. One very important drawback to the use of 'systems thinking' was identified by the historical geographer, Langton (1972). If we accept Hall and Fagan's definition, then it is evident that we can identify or create as many systems in thought or reality as we wish. This becomes, then, fundamentally an exercise in classification. It should then be clear that we need to be able to rate the value of different classifications and this, Langton contends, can only be done if we can identify the optimum outcome from the system we define. Whilst this may be feasible, say, in evaluating the water yield from a DRAINAGE BASIN in terms of irrigation, or as a site for a hydro-power operation, it is not at all clear that a slope, a PINGO or even the drainage basin can be legitimately thought to possess a 'goal'. This problem is, it seems, a fundamental one. And very closely linked to this inherent difficulty is that of identifying unique and non-overlapping systems in the real world. This is absolutely crucial for process-response systems if their operation is to be fully understood. Consider a drainage basin. Can its physical watershed be truly and exactly fixed? Can we be certain that the watersheds for mass or energy inputs and outputs coincide with the topographic limits? (Consider the question of subsurface water and solute movement.) If our designation of discrete basins may seem accurate, can we be as content with the identification of 'hillslopes' or even 'channels'? If we do not have this certainty – so different from the identification of (say) a household 'hot water system' – then it must surely cast doubt on the validity of our conclusions about the systems' operations. As a result, by the end of the twentieth century, the use of the term 'system' in geomorphology seemed really loosely, if at all, related to the specific ideas of von Bertalanffy, Strahler and Chorley. An example would be the text by Allen (1997) which opens with a chapter entitled 'Fundamentals of the Earth surface system' (pp. 1–50) with very little in the way to indicate direct descent from – say – Chorley and Kennedy, in terms of vocabulary or illustrations.

Before 1952, it was possible to think of 'systems' at all levels, both conceptual and practical. What might be termed the 'systems boom' of the 1950s–1970s in geomorphology argued for a situation in which a particular theoretical vision would inform the whole choice of research topics as well as the vocabulary in which results would be couched. In general the link between philosophical concept and practice was never really made. As a result, we have a residual terminology which really owes more to the vernacular than to either Strahler or Chorley. And we also see far more use of systems as a research tool in precisely those applied situations where a desired outcome can be specified, thus confirming Langton's assessment.

References

Allen, P.A. (1997) *Earth Surface Processes*, Oxford: Blackwell.

Bennett, R.J. and Chorley, R.J. (1978) *Environmental Systems: Philosophy, Analysis and Control*, London: Methuen.

Buffon, J.M.L. (1749) *Histoire naturelle, générale et particulière, avec la description du Cabinet du Roi*, Paris: L'Imprimerie Royale.

Chorley, R.J. (1960) Geomorphology and General Systems Theory, *US Geological Survey, Professional Paper* 500-B.

Chorley, R.J. and Kennedy, B.A. (1971) *Physical Geography: A Systems Approach*, London: Prentice Hall.

Hall, A.D. and Fagan, R.E. (1956) Definition of system, *General Systems Yearbook* 1, 18–28.

Langton, J. (1972) Potentialities and problems of adopting a systems approach to the study of change in human geography, *Progress in Geography* 4, 125–179.

Phillips, J.D. and Renwick, W.H. (eds) (1992) *Geomorphic Systems*, Amsterdam: Elsevier.

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.

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

Strahler, A.N. (1952) Dynamic basis of geomorphology, *Geological Society of America Bulletin* 63, 923–938.

von Bertalanffy, L. (1950) The theory of open systems in physics and biology, *Science* 3, 23–29.

SEE ALSO: Cycle of Erosion; drainage basin; flow regulation systems; integrated coastal management; land system; peneplain; step-pool system; threshold, geomorphic

BARBARA A. KENNEDY

T

TAFONI

Tafoni (singular tafone) are cavernous weathering forms which typically are several cubic metres in volume and have arch-shaped entrances, concave inner walls, overhanging margins (visors) and fairly smooth gently sloping, debris-covered floors (Mellor *et al.* 1997). They occur in many parts of the world (see Goudie and Viles 1997: table 6.1), including polar regions, but may be especially well developed in coastal and dryland situations. They occur in a wide range of rock types, but especially in medium and coarse-grained granites, sandstones and limestones. Indeed it is only rocks with relatively closely spaced discontinuities (bedding planes, foliation, joints) such as shales and slates, that seem to be relatively unaffected by these cavernous weathering forms.

The cavernous hollows of tafoni are believed to result largely from flaking and granular disintegration caused by a range of possible weathering processes that include hydration, salt crystallization, and chemical attack by saline solutions. Some workers have found clear evidence of SALT WEATHERING being involved, while others have not. The role of case hardening in their foundation is also the subject of debate, but can help to explain the formation of the visor. It is also possible that tafoni develop through a positive feedback effect, in that once a hollow is initiated it creates an environment in which weathering is favoured (Smith and McAlister 1986). For a cavity to grow there needs to be a mechanism to remove flakes and spalls. Wind may play a part, as may organisms such as pack rats. Although some early workers thought that the actual excavation of a cavity might be achieved by wind abrasion, many tafoni occur in environments where sand blasting does not occur or they may have an aspect (i.e. the leeward side of a boulder) or a height up a cliff face that precludes such a mechanism. It is clear that in whatever way they form, tafoni grow significantly over tens of thousands of years (Norwick and Dexter 2002).

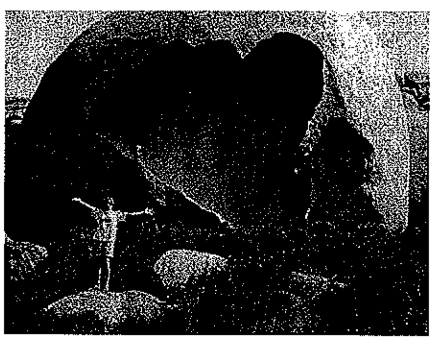


Plate 136 A large tafoni developed in granite near Calvi in Corsica

Plate 137 A remarkably developed tafoni in volcanic rocks in the Atacama Desert near Arica in northern Chile

References

Goudie, A.S. and Viles, H.A. (1997) *Salt Weathering Hazards*, Chichester: Wiley.
Mellor, A., Short, J. and Kirkby, S.J. (1997) Tafoni in the El Chorro area, Andalucia, southern Spain, *Earth Surface Processes and Landforms* 22, 817–833.
Norwick, S.A. and Dexter, L.R. (2002) Rates of development of tafoni in the Moenkopi and Kaibab formations in Meteor Crater and on the Colorado Plateau, northeastern Arizona, *Earth Surface Processes and Landforms* 27, 11–26.
Smith, B.J. and McAlister, J.J. (1986) Observations on the occurrence and origins of salt weathering phenomena near Lake Magadi, Southern Kenya, *Zeitschift für Geomorphologie NF* 30, 445–460.

A.S. GOUDIE

TALSAND

Large-scale infillings of ice-marginal valleys occurring in the Pleistocene lowlands of northern Germany and Poland. However, the term is also employed to represent any flat sandy region of glacifluvial and/or periglacial origin, set below the general level of Pleistocene uplands (Geest Plateau) (Schwan 1987). This is referred to as a talsand plain. Talsand is thus a geomorphological term that is largely restricted to north German Quaternary geology, translated as valley sand.

Talsand is composed of glacifluvial or periglacial sediments, though often an overlying bed of aeolian sands is present. The majority of talsand was deposited during the last glacial period (the Weichselian glacial *c.*70–10 ka years ago – the equivalent of the Devensian glacial in the UK, and the Wisconsin glacial in the USA), though formation may have initiated in the preceding late Saalian period. The overlying aeolian deposit typically originates from the late Weichselian and often continues into the Holocene. The transformation from fluvial to aeolian sediments is predominantly due to increased wind intensity and the lowering of ground water due to permafrost degradation (Schwan 1987).

Reference

Schwan, J. (1987) Sedimentological characteristics of a fluvial to aeolian succession in Weichselian Talsand in the Emsland (F.R.G.), *Sedimentary Geology* 52, 273–298.

STEVE WARD

TALUS

A term of French origin, which can have different meanings. For Anglo-Saxon geomorphologists it is used as a synonym for scree. It is an accumulation of weathered rock fragments under a cliff. The mechanisms which control this accumulation are, however, complex. The weathering of the underlying rock face is most often linked to frost (see FROST AND FROST WEATHERING) or earthquake action. The transit of rock debris from the cliff to the deposit can be due to one or several processes. Consequently, the morpho-sedimentological characteristics of this deposit depend on the type of these processes and the lithology.

In rare cases the accumulation is exclusively linked to rockfall. This rockfall talus presents a steep slope, a longitudinal sorting, and does not show any stratification in cross section. In most cases transit is due to a combination of different processes like avalanches or DEBRIS FLOWS. The morpho-sedimentological characteristics of the deposit will depend on the process or the dominant processes.

Further reading

Bertran, P. (ed.) (2003) *Dépôts de pente continentaux: dynamique et faciès*, BRGM.
Jomelli, V. and Francou, B. (2000) Comparing characteristics of rockfall talus and snow avalanche landforms in an alpine environment using a new methodological approach, *Geomorphology* 35, 181–192.

VINCENT JOMELLI

TALUVIUM

Taluvium is a slope deposit composed of rock fragments in a fine matrix, thereby bridging the gap between TALUS (composed of rock fragments) and colluvium (fine material only). Given time, talus will eventually change into taluvium, and subsequently into COLLUVIUM. The development of taluvium from talus has been related to weathering of the rock, although formation by the incorporation of finer airfall material in the talus lattice, such as by loess (Pierson 1982) and volcanic tephra, has also been suggested.

Taluvium deposits are stratigraphically, texturally and hydrologically heterogeneous, as talus formation and emplacement of the fine matrix is typically an episodic process. The accumulation of a fine matrix increases the total surface area of particle contact, increasing internal friction and encouraging aggregation. However, further accumulation of fines decreases the void ratio, impeding drainage and resulting in increasing pore-water pressures. Thus, the slope stability of taluvium (typically 25–28°) is predominantly influenced by its hydrological properties, with saturation resulting in a reduction in its stability angle to approximately half its dry value.

Reference

Pierson, T.C. (1982) Classification and hydrological characteristics of scree slope deposits in the northern Craigieburn Range, New Zealand, *Journal of Hydrology (New Zealand)* 21(1), 34–60.

STEVE WARD

TECTONIC ACTIVITY INDICES

Morphometric techniques used as reconnaissance tools to identify areas experiencing rapid tectonic deformation. Among the indices that have been found most useful are the following (Keller and Pinter 2002).

The hypsometric integral

The hypsometric curve describes the distribution of elevations across an area of land. It is created by plotting the proportion of total basin height (h/H = relative height) against the proportion of total basin area (a/A = relative area). The total height (H) is the relief within the basin (the maximum elevation minus the minimum elevation). The total surface area of the basin (A) is the sum of the areas between each pair of adjacent contour lines. The area (a) is the surface area within the basin above a given line of elevation (h). The value of relative area (a/A) always varies from 1.0 at the lowest point in the basin (where h/H = 0.0) to 0.0 at the highest point in the basin (where h/H = 1.0).

A simple way to characterize the shape of the hypsometric curve for a given drainage basin is to calculate its *hypsometric integral* (H_i). The integral is defined as the area under the hypsometric curve. One way to calculate the integral for a given curve is as follows [8,9]:

$$H_i = \frac{\text{Mean elevation} - \text{minimum elevation}}{\text{Maximum elevation} - \text{minimum elevation}}$$

A high hypsometric integral indicates a youthful topography. Digital Elevation Models (DEMs) make calculations easy (Gardner *et al.* 1990).

Drainage basin asymmetry

The Asymmetry Factor (AF) has been developed to detect tectonic tilting transverse to flow:

$$AF = 100\ (A_r/A_t)$$

where A_r is the area of the basin to the right (facing downstream of the trunk stream), and A_t is the total area of the drainage basin. AF values are sensitive to tilting perpendicular to the trend of the trunk stream and the greater the divergence of the AF values from 50 the greater is the degree of tilt.

Stream Length–Gradient Index (SL Index)

This is represented as:

$$SL = (\Delta H/\Delta L)L$$

where ΔH is the change in elevation of a stream reach and ΔL is the length of the reach. L is the total channel length from the midpoint of the reach upstream to the highest point on the channel. The index is used to identify recent tectonic activity by identifying anomalously high index values on a particular lithology.

Mountain-front sinuosity

This is an index that reflects the balance between erosional forces that tend to cut embayments into a mountain front and the tectonic forces that tend to produce a straight front coincident with an active range-bounding fault (see FAULT AND FAULT SCARP). It is expressed as:

$$S_{mf} = L_{mf}/L_s$$

where S_{mf} is the mountain-front sinuosity, L_{mf} is the length of the mountain front along the foot of the mountain at the pronounced break in slope, and L_s is the straight-line length of the mountain front. Given that mountain fronts associated with active tectonics and uplift are straight, they have low values of S_{mf}.

Ratio of valley floor width to valley height

This may be expressed as:

$$V_f = 2V_{fw}/[E_{ld} - E_{sc}) + (E_{rd} - E_{sc})]$$

Where V_f is the valley floor width-to-height ratio, V_{fw} is the width of the valley floor, E_{ld} and E_{rd} are elevations of the left and right valley divides respectively, and E_{sc} is the elevation of the valley floor. High values of V_f are associated with low uplift rates that have enabled streams to cut broad valley floors. Low values of V_f are associated with uplift.

The use of indices such as these, particularly in combination, enable the production of a relative tectonic activity class designation for an area.

References

Gardner, T.W., Sasowsky, K.C. and Day, R.L. (1990) Automated extraction of geomorphometric properties from digital elevation data, *Zeitschrift für Geomorphologie* 80, 57–68.
Keller, E.A. and Pinter, N. (2002) *Active Tectonics. Earthquakes, Uplift and Landscape*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.

A.S. GOUDIE

TECTONIC GEOMORPHOLOGY

As Burbank and Anderson (2001: 1) remark, 'The unrelenting competition between tectonic processes that tend to build topography and surface processes that tend to tear them down represents the core of tectonic geomorphology.' Immersed in small-scale process geomorphology, geomorphologists have been remarkably slow to explore both the significance of the conceptual advances provided by the PLATE TECTONICS model and the range of new techniques and data sources relevant to the quantification of long-term denudation rates (Summerfield 2000: 3). Nonetheless, over recent years the enormous growth in topographic data (see, for example, DIGITAL ELEVATION MODEL) and in geochronologic data (see, for example, COSMOGENIC DATING; FISSION TRACK ANALYSIS) have provided new opportunities to address long-standing questions of landscape evolution at the regional and continental scales (Morisawa and Hack 1985).

Plainly there are many Earth features that owe their form in large measure to tectonic activity (see ACTIVE AND CAPABLE FAULT; FAULT AND FAULT SCARP; FOLD; MANTLE PLUME; PULL-APART AND PIGGY-BACK BASIN; RING COMPLEX OR STRUCTURE; TECTONIC ACTIVITY INDICES). Equally there are miscellaneous types of tectonic activity (see CYMATOGENY; DIASTROPHISM; EPEIROGENY; ISOSTASY; SEAFLOOR SPREADING; WILSON CYCLE). At the large scale some major landscape features are associated intimately with various types of plate boundary (see ACTIVE MARGIN; ESCARPMENT; MOUNTAIN GEOMORPHOLOGY; PASSIVE MARGIN; SEISMOTECTONIC GEOMORPHOLOGY; VOLCANO, etc.)

The scope of modern tectonic geomorphology can be appreciated by the context of the recent text by Burbank and Anderson (2001). They start with a consideration of *geomorphic markers* (features or surfaces that provide a reference frame against which to gauge differential or absolute tectonic deformation). Next they consider dating methods, which are vital to determine rates at which faults move or surfaces deform. Then they pass on to stress, faults and folds, before analysing geodetic methods for analysing short-term deformation (e.g. GPS). This is followed by a consideration of the evidence for, and consequences of, palaeoseismology. Then they look at the balance between erosion and uplift, and deformation at various timescales from the Holocene to the Late Cenozoic. They conclude with a discussion of numerical modelling of landscape evolution.

References

Burbank, D.W. and Anderson, R.S. (2001) *Tectonic Geomorphology*, Oxford: Blackwell Science.
Morisawa, M. and Hack, J.T. (eds) (1985) *Tectonic Geomorphology*, Boston: Unwin Hyman.
Summerfield, M.A. (ed.) (2000) *Geomorphology and Global Tectonics*, Chichester: Wiley.

A.S. GOUDIE

TERMITES AND TERMITARIA

Termites, insects of which there are several thousand species, are members of the Isoptera order, and about four-fifths of the known species belong to the Termitidae family (Harris 1961). They vary in size according to their species, from the large African *Macrotermes*, with a length of around 20 mm and a wingspan of 90 mm, down to the Middle Eastern *Microcerotermes* which is only about 6 mm long with a wingspan of 12 mm.

They occur in great numbers – 2.3 million ha^{-1} in Senegal and 9.1 million ha^{-1} in the Ivory Coast

(UNESCO, UNEP, FAO 1979). The vast majority of termite species are found in the tropics, though their distribution is wider than this; they extend to 45–48°N and to 45°S.

Termite mounds and hills are the most striking manifestations of termite activity, and have a large range of sizes and morphologies (Goudie 1988). The heights of termite constructions vary considerably according to species. There are records in the literature of mounds attaining heights in excess of 9 m, though most are less than this. Among the species that create the tallest mounds are *Bellicositermes bellicosus*.

In general the densities of mounds vary considerably according to both environmental conditions (e.g. soil type) and termite species. So, for example, the density of the very large mounds produced by *Macrotermes bellicosus*, *Macrotermes subhyalinus*, *Macrotermes falciger*, *Bellicositermes bellicosus* and *Nasutitermes triodiae* tends to be less (often around 2–10 ha^{-1}) than those for the smaller types of mound (often around 200–1,000 ha^{-1}).

Undoubtedly soil characteristics are an important control of mound formation. Mounds tend to be rare on sands (where there is insufficient binding material), on deeply cracking self-mulching clays (which are unstable), or on shallow soils

Plate 138 A large termite mound developed by *Macrotermes* in the Mopane woodland of northern Botswana

(where there is a shortage of building material) (Lee and Wood 1971). Soil drainage may also be important, as are human activities. Although in most cases it is obvious that particular mounds have been produced by particular species of termites there have been some arguments about the origin of some mounds found in the tropics and elsewhere (see MIMA MOUND). For example, Cox and Gakaha (1983) have argued that certain mounds in Kenya are created by the mole-rat, *Tachyoryctes splendens*, whereas Darlington (1985) argues that the same mounds are produced by a type of termite – *Odontotermes*.

Whereas termites live in hidden subsoil chambers or in the conspicuous mounds just discussed, they have a significant effect on soils, partly because of their mechanical activities and partly because of their feeding habits. They can, for example, cause soils to become rich in calcium, play a major role in nutrient cycling, remove organic litter from soil surfaces and mobilize particular soil fractions (e.g. clay). Whether they contribute to laterite formation or lead to its degradation, however, has been the subject of debate (Runge and Lammers 2001).

Termites can contribute to accelerated rates of soil denudation. Lee and Wood (1971) identify three main ways in which termites can do this:

1 by removing the plant cover;
2 by digesting or removing organic matter which would otherwise be incorporated into the soil, and thus making the soil more susceptible to erosion;
3 by bringing to the surface fine-grained materials for subsequent wash and creep action.

The huge numbers of termites and their large total biomass in favoured localities ensures that these three mechanisms are important. The live weight biomass of termites can be comparable to the live weight biomass of large mammalian herbivores in tropical areas. However, in addition to the potential for erosion and sediment yield caused by mound formation and abandonment it is important to remember the other major consequence of termite-caused soil translocation. This is the construction of covered runways or 'sheetings'. These are constructed of soil particles cemented together with salivary secretions (Bagine 1984).

Through their effects on denudation termites may have more widespread effects on fluvial systems. Drummond (1888: 158) postulated that while

Egypt was the gift of the Nile, that river's sediments resulted from 'the labours of the humble termites in the forest slopes about Victoria Nyanza'.

References

Bagine, R.K.N. (1984) Soil translocation by termites of the germs Odontotermes (Holmgren) (Isoptera: Macrotermininae) in an arid area of northern Kenya, *Oecologia* 64, 263–266.
Cox, G.W. and Gakaha, C.G. (1983) Mima mounds in the Kenya Highlands: significance of the Dalquest-Scheffer hypothesis, *Oecologica* 57, 170–174.
Darlington, J.P.E.C. (1985) The underground passages and storage pits used in foraging by a nest of termite Macrotermes michaelseni in Kajiado, Kenya, *Journal of Zoology*, London 198, 237–247.
Drummond, H. (1888) *Tropical Africa*, London: Hodder and Stoughton.
Goudie, A.S. (1988) The geomorphological role of termites and earthworms in the tropics, in H.A. Viles (ed.) *Biogeomorphology*, 166–197, Oxford: Blackwell.
Harris, W.V. (1961) *Termites: Their Recognition and Control*, London: Longman.
Lee, K.E. and Wood, T.G. (1971) *Termites and Soils*, London and New York: Academic Press.
Runge, J. and Lammers, K. (2001) Bioturbation by termites and Late Quaternary landscape evolution in the Mbomou Plateau of the Central African Republic, *Palaeoecology of Africa* 27, 153–169.
UNESCO, UNEP, FAO (1979) *Tropical Grazing Land Ecosystems*, Paris: UNESCO.

A.S. GOUDIE

TERRACE, RIVER

A river terrace is the planar surface that remains after the river, which formed it, incised its former valley floor. River terraces are abandoned river channels and floodplains. Their presence in river valleys throughout the world provides a record of changes in the flow regimes of rivers and the sediment supplied to them over time.

The flat surface of a terrace, called the *tread*, represents the highest elevation of the valley floor before incision occurred. Terrace treads, also called benches or platforms, are composed of alluvium, bedrock, or bedrock covered with a thin deposit of alluvium. They dip downvalley, recording the gradient of the channel that formed them, unless tectonic or isostatic uplift has subsequently altered their slope. The steep slope between treads or between the tread and the active floodplain is called the *riser*. Flights of terraces represent multiple, discrete episodes of downcutting, punctuated by periods of stability or AGGRADATION. Terraces may be continuous along a valley, or discontinuous if portions of the same terrace have become separated by tributary entrenchment or other geomorphic processes. More recent deposits, including those of mass movements, alluvial fans, volcanic ash, or wind-blown fines, may bury terrace treads. In a tectonically active area, faulting can alter the height relationships between terraces.

Genetically, terraces are considered to be either depositional (fill) or erosional (cut) landforms. Depositional terraces form as a result of the aggradation and later entrenchment of alluvium. They are abandoned floodplains, and stratigraphically show vertical and lateral processes of sediment accretion. Erosional terraces are surfaces formed by the erosional removal of bedrock or alluvial fill from the former valley floor. The term river terrace is used inclusively to describe the landform without specifying the materials (bedrock or fill) or the genetic processes (erosional or depositional) responsible for a specific feature. When the ages or relative ages of a flight of terraces are known, terrace surfaces are usually identified numerically, with the lowest number (1) used for the oldest surface. By recording changes in the flow regime of rivers, terraces are important indicators of tectonic, climatic, and even anthropogenic environmental history.

Terrace-forming processes

Different processes and circumstances, acting alone or in combination, cause rivers to incise, stranding their former valley floors above the active channel. Incision may begin gradually or catastrophically. A river cuts down through its own deposits when greater flow energy, lower BASE LEVEL, and/or less sediment load increase its erosional capacity. Discharge and stream energy can be increased by changes in climate and by processes, including base-level lowering, that steepen the channel gradient. Climatically driven incision occurs when climate becomes wetter, when ice melts (warmer climate), or when upstream climate–vegetation–soil relationships lead to conditions of flashier rainfall runoff. The latter would occur where decreasing precipitation causes a marginally semi-arid area to become more arid such that vegetation becomes sparser, soil erosion accelerates, infiltration rates decrease and the proportion of rainfall flowing to the river as storm runoff thereby increases. Anthropogenic activities, including forest removal and road building, also increase the flashiness of runoff. Factors that cause the channel gradient to steepen increase the energy of the flow, allowing the river to entrain materials previously deposited. Increased gradients result from tectonic and isostatic uplift, headward propagation of knickpoints, faulting, or lowering the erosional base level. Many of the terraces in contemporary landscapes are Pleistocene or Holocene in age and record changes in river regimes due to shifts in climate and in geomorphic process regimes between glacial and interglacial periods. Among the multiple factors leading to Pleistocene river incision and terrace formation was the lowering of sea level during glacial periods, when more water was stored on land as ice. Glacial period sea levels dropped more than 100 m, lowering erosional base levels and steepening continental river channel gradients accordingly (see GRADE, CONCEPT OF).

Changes in sediment supply tip the balance between aggradation and degradation in a river. A river's capacity to transport sediment derives from the volume and calibre of sediment supplied to it and the energy available to move the sediment. In a period of little environmental change, the channel geometry of a river, including its gradient, represents the discharge, energy and sediments characteristic of that environment at that time. When active glaciers, for example, provide an abundant sediment supply, the resulting channel will become steep, swift, shallow and probably braided. Reducing the sediment supply leaves such a river, adapted for heavy sediment loads, with excess energy. Thus, reducing the sediment supply upstream leads to more aggressive erosion and downcutting downstream, a condition that will continue until the flow regime becomes more in balance with the available energy and sediment. The supply of sediment to a river can be diminished by changes in climate or land management practices that increase vegetative cover or decrease mass movements and wind erosion on upstream land surfaces. Dams stop sediment, too. Such changes increase the erosional energy of the river downstream, possibly to the point that it will begin to incise its own deposits. RIVER CAPTURE of a higher gradient, sediment-laden headwater river by a lower gradient piedmont river causes the piedmont river to gain additional sediment input from the high gradient headwaters, and the headwater river to adjust to a lower local base level. Steepening gradients in the capturing stream can cause flow energy to increase, and the discharge of the capturing stream may also increase due to the increased size of its drainage basin. Finally, catastrophic events, e.g. outburst flooding from a glacial lake or a landslide-dammed lake, can trigger catastrophic incision and initiate terrace formation.

Depositional terraces

A depositional (fill, aggradational) terrace is a former floodplain that has been incised by the river. It differs from an active floodplain by being too far above the river channel to convey the overbank flows of the mean annual flood. Massive alluvial deposits in large terraces represent conditions of abundant sediment supply to the river (Figure 164a). The sediments in depositional terraces were built up by the accretion of alluvial sediment, either vertically or laterally, during a period in which that surface was the active floodplain of the river. For floodplains built by vertical accretion, a cross section of the terrace would reveal horizontal stratification, with flood deposits on top of flood deposits sorted by size and fining upward. Gravels in the deposits would be rounded and probably imbricated and aligned to point downstream. If the floodplain had built up by lateral accretion, the deposits would have the stratigraphy of POINT BARS. Depositional terraces may be metres to kilometres wide, and follow a present or former course of the river for thousands of kilometres, often on both sides of the channel. Flights of depositional terraces occur where rivers have developed new floodplains between separate episodes of downcutting.

In many examples around the world, repeated episodes of valley filling followed by major evacuation of sediment have created younger (inner) depositional terraces developed on fill material which was emplaced after the valley was scoured out between higher, older terraces (Figure 164b). Depositional terraces can be distinguished from erosional terraces in that (1) the tread surface of a depositional terrace represents the uneroded surface of a valley fill and (2) the underlying rock surface (if its topography can be determined seismically or viewed in a transverse cut) may be very irregular.

Erosional terraces

The surface of an erosional river terrace was levelled by lateral fluvial erosion before the river

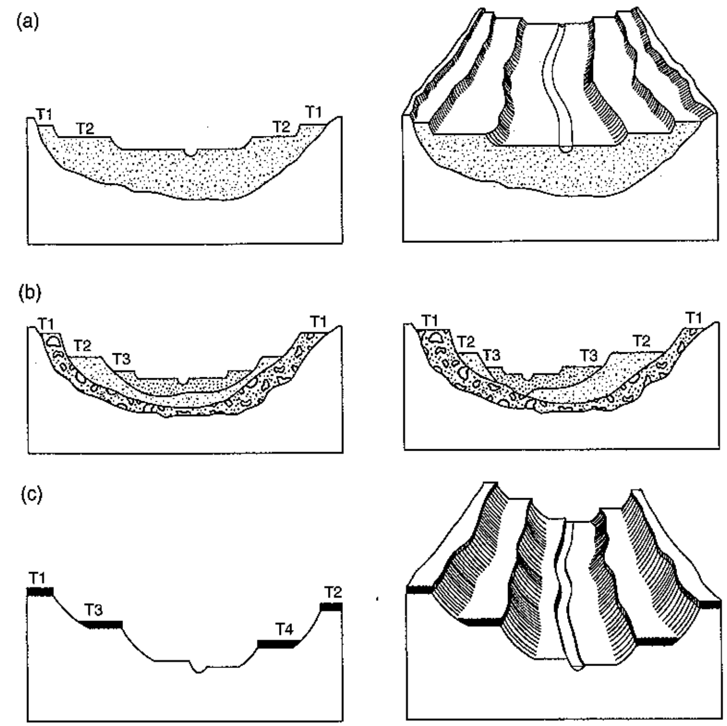


Figure 164 Valley cross section and block diagrams illustrating (a) paired depositional terraces, (b) depositional terraces formed from multiple episodes of valley fill and entrenchment, and (c) unpaired strath terraces

incised to form the terrace. Like depositional terraces, erosional terraces are remnants of an older valley floor, into which the river has cut down. Erosional terraces scoured from bedrock are called *bench*, *strath*, or *rock-cut* terraces (Howard *et al.* 1968: 1,117). The term 'strath' is a Scottish word meaning wide valley. Rivers carve straths by lateral corrasion, with abrasive alluvial sediments eroding the channel bed from side to side as meanders migrate downstream and/or the meander radius increases. The underlying bedrock surface, as nearly planar as channel beds and valley bottoms are today, lies parallel to the surface of the thin cap of alluvium (Figure 164c). The thickness of the alluvial cap indicates the depth of scouring in the former river regime: the alluvium must be thin enough so that the river could have been in erosional contact with the bedrock valley bottom. A thick valley fill, on the other hand, protects the valley bottom from erosion. The maximum thicknesses of alluvial lag deposits, usually gravel, on strath terraces depends on the size and energy level of the river. Mackin (1937: 828) reported 2.5 m of gravel with a cover of silt in the Shoshone Valley (Wyoming, USA); Wegmann and Pazzaglia (2002: 734) suggest 3 m as a typical maximum depth.

Not all erosional terraces are strath terraces with eroded bedrock at or just below the surface. A valley floor composed of unconsolidated fill may also become truncated by lateral erosion and subsequently stranded by incision to become a river terrace. Such terraces are called *fill*-cut (Bull 1990: 355) or *fill-strath* (Howard *et al.*

1968: 1,119). *Structural* terraces are erosional benches of resistant bedrock resulting from differential erosion rather than from changes in the river flow regime. Structural terraces abound in the Grand Canyon of the Colorado River, where contrasting erodibilities of the nearly horizontally bedded sedimentary strata cause the resulting canyon walls to appear as rock steps.

Paired and unpaired river terraces

Where terraces occur at the same elevation across a valley, they are called *paired* (Figure 164a); otherwise they are *unpaired*. Paired and unpaired terraces can be erosional or depositional; they can be formed on bedrock or on alluvial fill. Paired terraces represent conditions in which downcutting predominates over lateral cutting. A river flowing across the middle of a thick floodplain would respond to a drop in base level by incision. The result would be a lower channel between a pair of terraces. If the base level remained at the same low position for an extended period of time, the period of downcutting, accompanied by mass movements of channel banks, would be followed by a period of aggradation in which the river might form a new floodplain along the entrenched channel. A second episodic event that renewed the greater erosional capability of the river could initiate a second phase of downcutting, causing the river to abandon the newer floodplain and incise a second, inner set of paired terraces. Such pairs of depositional terraces are assumed to be of equal age.

Unpaired river terraces reflect conditions in which downcutting is slow and lateral erosion is occurring at the same time (Ritter 1986: 269). Lateral migration of the river channel can erode and eliminate some older terraces, leaving an asymmetrical record of depositional or erosional surfaces. Unpaired terraces are unlikely to be of equal age, and likely to be erosional in origin. Strath terraces are typically unpaired.

Terraces as evidence of environmental change

Terraces may record tectonic events, changes in climate and other environmental changes that alter the erosional capacity and sediment load of a river. Dating terrace surfaces enables researchers to calculate incision rates and better understand the climate or tectonic history of a region. The age of a terrace is determined by its relative position in the valley and by other evidence present in its biological, chemical and anthropogenic record. Alluvial particles on a strath terrace are synchronous with the time period of erosion, whereas alluvial particles in a depositional terrace predate the time of terrace formation by the time required for a slug of sediment to travel to that location in the river system and the time lapse between deposition and incision (Bull 1990: 360). Organic matter buried in the terrace can be dated with radiocarbon (^{14}C) dating techniques (e.g. Wegmann and Pazzaglia 2002: 734), and investigators have begun to date clasts on terrace surfaces based on the concentration of cosmogenic isotopes such as ^{10}Be and ^{26}Al (e.g. Hancock *et al.* 1999: 47) (see DATING METHODS). Terrace ages can sometimes be inferred from the environmental setting portrayed in the biological record. The presence of sub-Arctic molluscs and fossil ice-wedge casts in terraces along the River Thames indicates deposition of terrace material during cold periods, interpreted as glacial periods in the Pleistocene (Goudie 1984: 292). Terraces from the late Tertiary or Pleistocene can be distinguished from Holocene terraces, not only by their biota and their position in the landscape, but by the degree of weathering of terrace materials. In some older terraces, formerly unconsolidated terrace materials have become cemented by carbonate, silica or iron oxides (Costa and Baker 1981: 161). Human artefacts, including ruined structures and Roman coins (Judson 1963: 899), have helped date Late Holocene terrace surfaces.

In some investigations, terrace formation is attributed to a single causative factor. Born and Ritter (1970: 1,240), for example, attributed the flight of terraces in the Truckee River above Pyramid Lake (California, USA) to an anthropogenically lowered base level. Alternatively, terrace formation can represent a complex landscape response to one change (e.g. a climatic factor) or the integrated landscape response to a set of changes (e.g. climatic and tectonic). The terrace record of the River Rhine, which records the climatic and erosional history of the region, has also been affected by uplifts in the middle and subsidence in the lower portions of the valley (Fairbridge 1968: 1,131).

Evidence provided by river terraces can be complex and challenging to interpret. Erosional episodes can remove older terraces; in fact, terrace sequences are rarely found completely intact. Terraces in one drainage basin may respond more to local factors than to regional climatic or tectonic controls. As an example of this, Brakenridge (1981: 75) found terrace formation along the Pomme de Terre River in southern Missouri (USA) to not match either the sea-level history for the Gulf of Mexico or the glacial chronology of the upper Missouri–Mississippi basin. In other examples, Ritter (1982: 352) found terraces in one valley in the Alaska Range to have formed after a moraine-dammed lake overtopped a drainage divide, and Wegmann and Pazzaglia (2002: 740) did not find base-level control responsible for terrace formation in the Clearwater River (Olympic Mountains, USA), even though the river flows directly into the Pacific Ocean. They suggest that the Clearwater River operates at or near its capacity for sediment transport and responds to changes in upstream sediment production, some of which is likely to be earthquake-related, with episodes of vertical (terrace-forming) or lateral (valley-widening) incision. Their interpretation is supported by Schumm's (1975: 77) work showing COMPLEX RESPONSE of a fluvial system could lead to small terrace formation without external forcing variables. Bull (1990: 352) emphasizes a scale difference between major terraces, particularly climatically caused aggradation surfaces and large tectonically caused straths, and minor terraces, which develop in response to local factors.

Although terrace tread formation is generally thought to reflect relatively stable conditions over long periods of time, terraces are also known to have formed catastrophically or over intervals of only a few years. From photographic and historical evidence, Born and Ritter (1970: 1,240) documented the formation of at least six well-developed river terraces upstream of Pyramid Lake (California, USA) in the forty-four years since water diversion began to lower the lake level.

Tread surfaces of terraces in the contemporary landscape may not have been disturbed by erosion or deposition since the time of their abandonment by the active river channel. A flight of such terraces presents a CHRONOSEQUENCE of weathering and soils, and terraces of known age present special opportunities for studying soil development (Bull 1990: 352). In inhabited areas, opportunities for the study of terrace soils are commonly constrained by human disturbance. Younger terraces are sought as sources of sand and gravel, and terrace surfaces are often chosen for agriculture, urbanization, and the location of highways and airports. Their relative flatness in high-relief environments and their position above elevations of frequent flooding and poor drainage makes terrace surfaces attractive for human occupation.

References

Born, S.M. and Ritter, D.F. (1970) Modern terrace development near Pyramid Lake, Nevada, and its geologic implications, *Geological Society of America Bulletin* 81, 1,233–1,242.

Brakenridge, G.R. (1981) Late Quaternary floodplain development along the Pomme de Terre River, southern Missouri, *Quaternary Research* 15, 62–76.

Bull, W.B. (1990) Stream-terrace genesis: implications for soil development, *Geomorphology* 3, 351–367.

Costa, J.E. and Baker, V.R. (1981) *Surficial Geology. Building with the Earth*, New York: Wiley.

Fairbridge, R.W. (1968) Terraces, fluvial–environmental controls, in R.W. Fairbridge (ed.) *Encyclopedia of Geomorphology*, 1,124–1,138, New York: Reinhold.

Goudie, A. (1984) *The Nature of the Environment*, Oxford: Basil Blackwell.

Hancock, G.S., Anderson, R., Chadwick, O. and Finkel, R. (1999) Dating fluvial terraces with ^{10}Be and ^{26}Al profiles: application to the Wind River, Wyoming, *Geomorphology* 27, 41–60.

Howard, Arthur D., Fairbridge, R.W. and Quinn, J.H. (1968) Terraces, fluvial – Introduction, in R.W. Fairbridge (ed.) *Encyclopedia of Geomorphology*, 1,117–1,123, New York: Reinhold.

Judson, S. (1963) Erosion and deposition of Italian stream valleys during historic time, *Science* 140, 898–899.

Mackin, J.H. (1937) Erosional history of the big Horn Basin, Wyoming, *Geological Society of America Bulletin* 48, 813–893.

Ritter, D.F. (1982) Complex terrace development in the Nenana Valley near Healy, Alaska, *Geological Society of America Bulletin* 93, 346–356.

——(1986) *Process Geomorphology*, 2nd edition, Dubuque, IA: William C. Brown.

Schumm, S.A. (1975) Episodic erosion: a modification of the geomorphic cycle, in W. Melhorn and R. Flemal (eds) *Theories of Landform Development*, 69–85, London: George Allen and Unwin.

Wegmann, K. and Pazzaglia, F. (2002) Holocene strath terraces, climate change, and active tectonics: the Clearwater River basin, Olympic Peninsula, Washington State, *Geological Society of America Bulletin* 114, 731–744.

Further reading

Selby, M.J. (1985) *Earth's Changing Surface*, Oxford: Oxford University Press.

SEE ALSO: floodplain; fluvial geomorphology; sediment load and yield; valley

CAROL HARDEN

TERRACETTE

A miniature unvegetated step-like feature that forms on hillslopes. Terracettes extend across the slope in a parallel manner, though predominantly following the contours of the land. They commonly form on ground that is fairly unconsolidated, particularly pasture, possessing moderate to steep hillslope gradients. Terracettes are rarely greater than 0.5 m in height and depth, with a spacing of about 1 m, and may extend laterally for tens of metres. They may form in a variety of climatic environments. The vertical drop exhibited in a terracette is known as the riser whereas the horizontal platform is called the tread. Additionally, the base of the riser is referred to as the foot of the terracette, while the point where the tread meets the riser is termed the crown.

Terracettes are irregular and often anastomosing forms, and may feature as intermittent steps or a whole network of steps covering a hillside. On a loessic slope of 24°, it has been estimated that terracettes may form on up to 11 per cent of the ground surface, and as much as 40 per cent on a slope angle of 37° (Selby 1993). Further terms for a terracette include pseudo-terraces or false terrace. Angle of slope is crucial for terracette formation. An average slope angle of about 30° dictates the boundary for terracette formation. Below this angle the slope face will not break, but instead will exhibit a distinctive undulating surface. Surficial material (based on grass) will begin to crack at angles greater than 30°, while at angles greater than 50° cracks occur at the back of each tread and small-scale slumping occurs, thus developing the characteristic step-like feature (Selby 1993: 258).

The origin of terracettes is contentious, and several explanations of their development have been produced. However, it is likely that the following mechanisms of formation are interrelated, and that each example is the result of a mix of mechanisms, unique to the site. The dominant mechanism of terracette formations is by soil creep, and occurs when hillslope angles are greater than that of which the unconsolidated mantle material can remain stable, resulting in slippage. Terracette formation is not limited to open grassland, and may develop in forested land as soil is washed downslope and leaf litter accumulates in the wake of tree roots. Additionally, erosion may be provoked or enhanced by the trampling of land by livestock. The hooves of both cattle and sheep can remove and wear down the surficial materials, particularly upon frequently used tracks and walkways (termed cattle or sheep tracks).

Terracettes can also result from near-surface faulting, which may break the overlying land into a series of small terracettes. Additionally, they can also be ablation products, man-made aids for cultivation on hillslopes (lynchets), as products of solifluction and antiplanation, or as miniature fluvial terraces.

Reference

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

Further reading

Rahm, D.A. (1962) The terracette problem, *Northwest Science* 36, 65–80.
Vincent, P.J. and Clarke, J.V. (1976) The terracette enigma – a review, *Biuletyn Peryglacjalny* 25, 65–77.

STEVE WARD

TERRAIN EVALUATION

The term *terrain evaluation* has been used to describe a wide range of geomorphological techniques and no single definitive meaning has been established. In its narrowest definition terrain evaluation is regarded as synonymous with mapping of LAND SYSTEMS, a procedure for classifying the landscape by dividing it into landform assemblages with similarities in terrain, soils, vegetation and geology (Mitchell 1973). In a slightly broader definition, Lawrance *et al.* (1993) regarded terrain evaluation for engineering projects as a method for summarizing the physical aspects of a landscape initially through classification and then including an assessment of ground conditions in terms of engineering requirements. Griffiths and Edwards (2001) use the expression *land surface evaluation* as an alternative to terrain evaluation in ENGINEERING GEOMORPHOLOGY and engineering geology because the varied usage had created confusion and led to misunderstanding. However, in practice terrain evaluation and land surface evaluation are synonymous and therefore the definition proposed by Griffiths and Edwards (2001) is the most appropriate to use. This states that it is the evaluation and interpretation of land surface and near-surface features using techniques that do

not involve ground exploration by excavation (except using small hand-dug pits or hand auger holes) or geophysics. Based on this definition terrain evaluation can be regarded as integral to the development of the ground model proposed by Fookes (1997) as central to all successful civil engineering construction. The definition would also be suitable when terrain evaluation is used as a technique of APPLIED GEOMORPHOLOGY in planning and environmental studies (see Smith and Ellison 1999). In these studies, the terrain evaluation procedure would include an evaluation of soils, vegetation, land use, materials, drainage and human activity in addition to geomorphological processes and landforms.

The techniques that can be employed in terrain evaluations include: GEOMORPHOLOGICAL MAPPING, geological mapping, engineering geological mapping, remote sensing interpretation, analytical photogrammetry, land systems mapping, natural hazard and risk assessment, and the use of Geographical Information Systems (GIS). The output from a terrain evaluation is usually a suite of maps either held as hardcopy or as a suite of overlays in a GIS. The map data can be classified into three categories:

1 Factual or element maps that record the actual ground conditions (Table 45a).
2 Derivative maps that are obtained by either combining element maps or are based on an interpretation of the element maps (Table 45b).
3 Summary maps that pull together a range of derivative and element maps to identify combinations of hazards, resources or land use issues that act either as constraints to any development or indicate the potential of the land for exploitation (Table 45c).

Table 45 Terrain evaluation map categories

Terrain evaluation map category	Examples of typical maps
(a) Element maps	• morphology • topography • bedrock geology • superficial geology • lithology • vegetation • pedology • land use • geotechnical properties • location of sites of special scientific interest • exploratory holes and wells • hydrology
(b) Derivative maps	• slope steepness that uses topography to classify maps into distinct groups based on morphology and topography • depth to bedrock that utilize data from the geological maps • geomorphology • hydrogeology • depiction of various types of resources, such as sand and gravel or brick clay • foundation conditions for engineering structures • geotechnical zoning, i.e. areas of homogeneous ground conditions • hazards, such as subsidence, landslides, flooding or contaminated land • previous industrial usage
(c) Summary maps	• development potential • potential resources • planning constraints, including statutory protected land • construction constraints

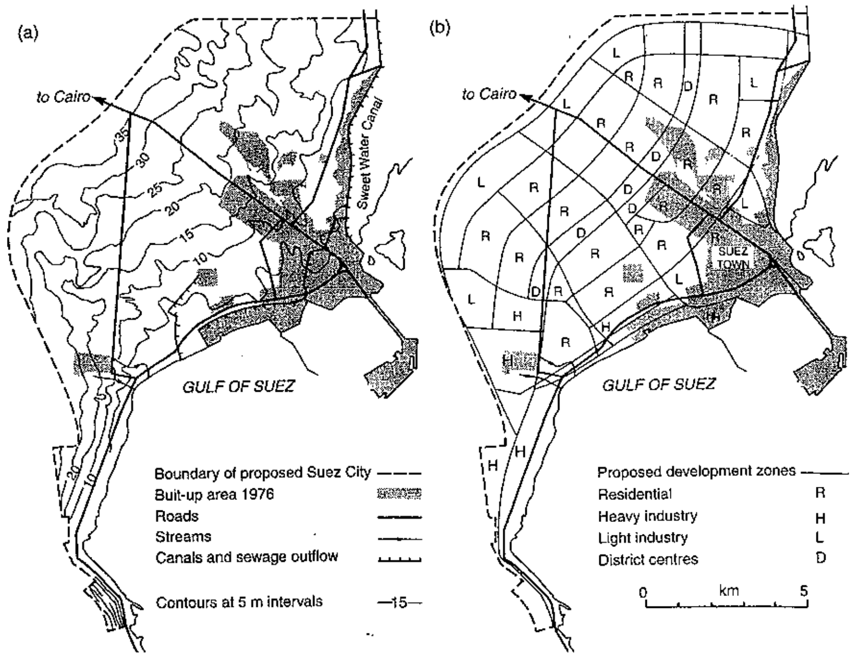


Figure 165 Maps of Suez, Egypt: (a) topography and extent of urban area in 1976; (b) proposed layout of an enlarged urban area (after Jones 2001, reprinted with permission of the Geological Society of London)

An extended legend is normally attached to each of the maps and most terrain evaluation studies would include an interpretative report that explains the basis for the development of the derivative and summary maps.

Terrain evaluation is best illustrated through a case study and Jones (2001) provides a classic example of its use for flood hazard assessment. In this study of a proposed development for Suez New City (Figure 165) geomorphological mapping was initially undertaken based on aerial photograph interpretation and field mapping. This resulted in the production of a geomorphological map, originally at a scale of 1:25,000, that established the distribution of a suite of marine, fluvial and bedrock features in addition to existing areas of urban development (Figure 166). These data were then combined with an evaluation of WADI catchment areas and channel form, based on further aerial photograph interpretation and field data analysis, to produce an interpretative map of flood hazard utilizing an ordinal scaling system (Figure 167). The final hazard map provided the base for both planning development and identifying the areas requiring flood protection.

References

Fookes, P.G. (1997) Geology for engineers: the geological model, prediction and performance, *Quarterly Journal of Engineering Geology* 30, 290–424.

Griffiths, J.S. and Edwards, R.J.G. (2001) The development of land surface evaluation for engineering practice, in J.S. Griffiths (ed.) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication No. 18, 3–9.

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 No. 18, 159–170.

Lawrance, C.J., Byard, R.J. and Beaven, P.J. (1993) *Terrain Evaluation Manual*, Transport Research

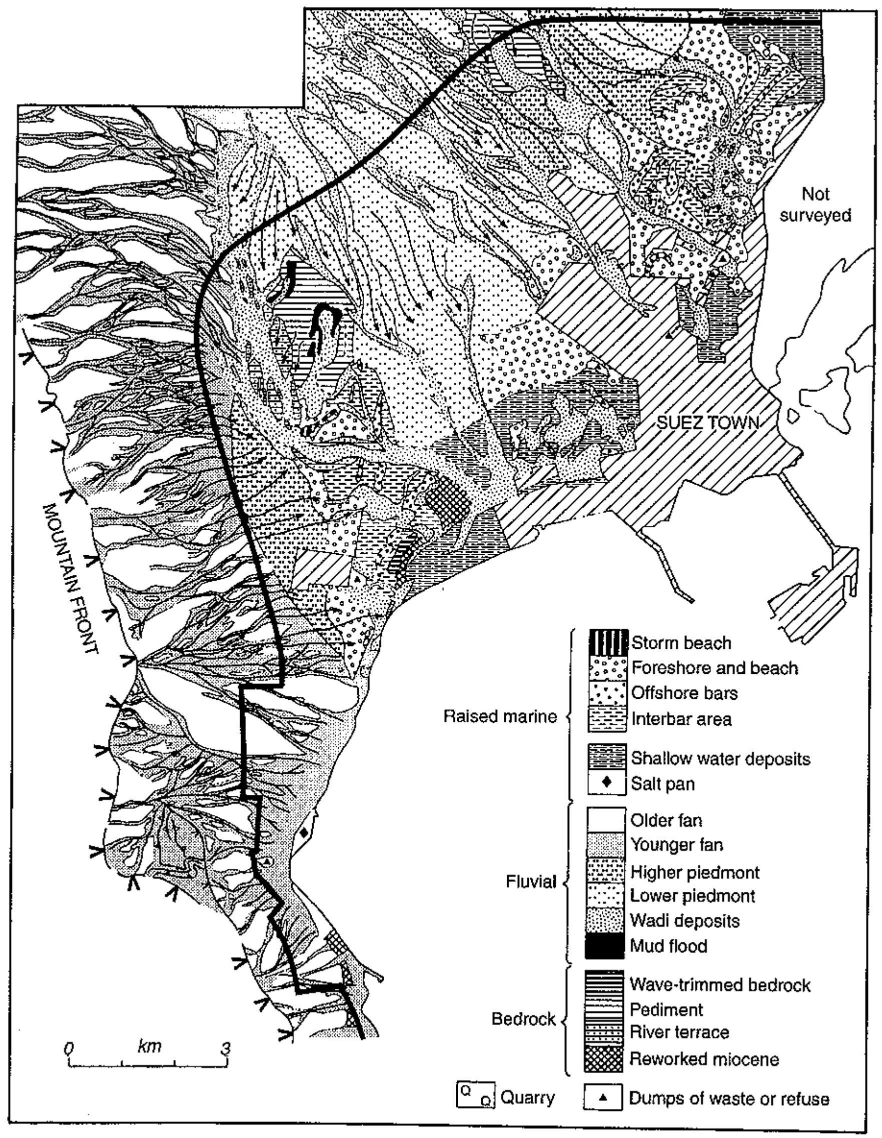


Figure 166 Geomorphological map of the Suez area produced by aerial photograph interpretation and ground mapping (after Jones 2001, reprinted with permission from the Geological Society of London)

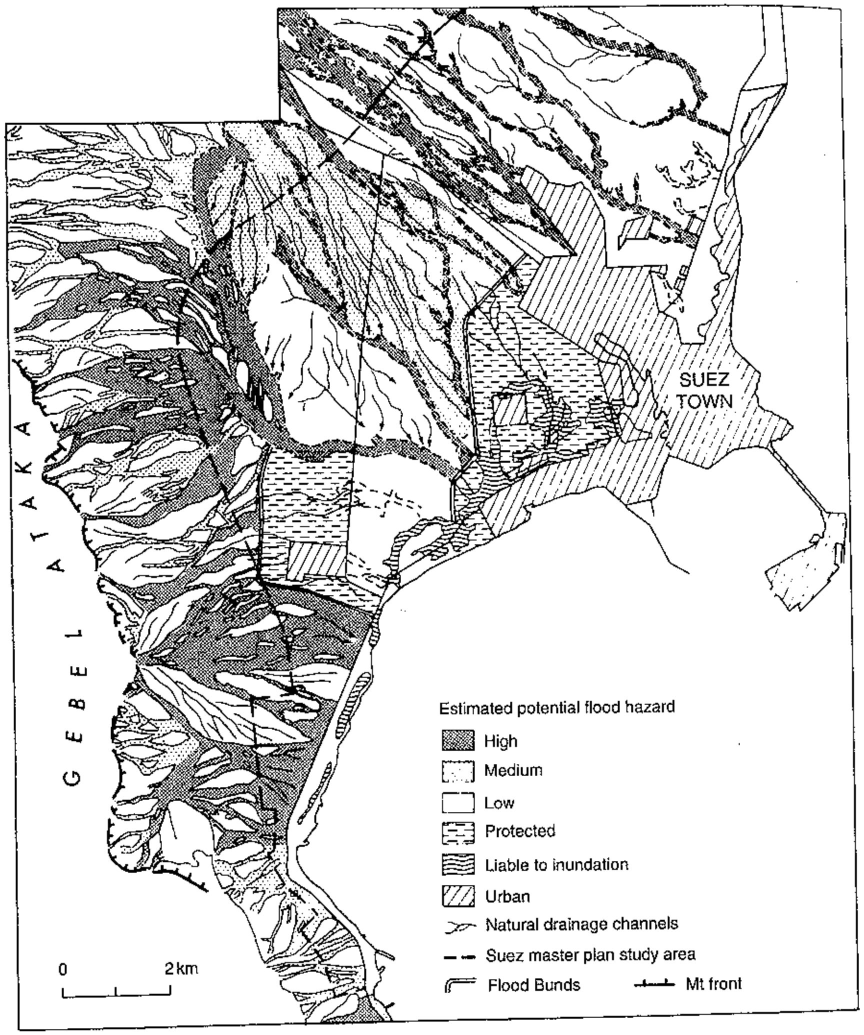


Figure 167 Flood hazard map of the Suez area (after Jones 2001, reprinted with permission from the Geological Society of London)

Laboratory, State of the Art Review 7, London: HMSO.
Mitchell, C.W. (1973) *Terrain Evaluation*, London: Longmans.
Smith, A. and Ellison, R.A. (1999) Applied geological maps for planning and development: a review of examples from England and Wales 1983 to 1996, *Quarterly Journal of Engineering Geology* 32, S1–S44.

Further reading

Cooke, R.U. and Doornkamp, J.C. (1990) *Geomorphology in Environmental Management*, 2nd edition, Oxford: Oxford University Press.
Edwards, R.J.G. (2001) Creation of functional ground models in an urban area, in J.S. Griffiths (ed.) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication No. 18, 107–113.
Fookes, P.G., Lee, E.M. and Sweeney, M. (2001) Pipeline route selection and ground characterization, Algeria, in J.S. Griffiths (ed.) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication No. 18, 115–121.
Griffiths, J.S. (ed.) (2001) *Land Surface Evaluation for Engineering Practice*, Geological Society Engineering Geology Special Publication No. 18.
Waller, A.M and Phipps, P. (1996) Terrain systems mapping and geomorphological studies for the Channel Tunnel rail link, in C. Craig (ed.) *Advances in Site Investigation Practice*, 25–38, London: Thomas Telford.

JAMES S. GRIFFITHS

THERMOKARST

The term 'thermokarst' was created in 1932 by a Russian, Yermolayev (Czudek and Demek 1970: 103), to describe an uneven morphology (thermokarst terrain) with some similarities to karstic morphology and resulting from soil subsidence due to the melting of ice within PERMAFROST. Later, the same word was used mainly to refer to the processes of soil subsidence resulting from thaw.

A prerequisite to thermokarstic terrain is excess ice within permafrost; in other words, the volume of ice must exceed the volume of the pores that water, under natural conditions, can fill in the soil. If excess ice is present, some change, even a small one, in the surface conditions (e.g. vegetation) can induce the melting of the upper part of the permafrost; such melting recurs during several summers and causes collapse that, cumulatively, can be important. Such collapses will continue until the ACTIVE LAYER has regained a sufficient thickness to protect the permafrost from further melting. Excess ice does not occur everywhere in permafrost. It is mainly found in the low ground of valley bottoms and in coastal lowlands containing abundant silty clays – locations of thaw-sensitive permafrost. Thermokarst also refers to collapse resulting from the disappearance of glacial ice. The forms resulting from melting of ice differ according to the distribution of the ice in the soil and, thus, according to the types of ice. The principal forms of thermokarst are summarized below.

Thermokarst lakes are widespread in the coastal lowlands of Siberia, Alaska and the Mackenzie Delta area (Canada), where excess ice is abundant in the soil. Such lakes are rounded depressions and are rather shallow. They progressively enlarge and their diameter can reach 1–2 km.

In some regions, thermokarst lakes are elongated and under influence of the direction of prevailing winds. They are called 'oriented lakes'. The mechanism leading to such a shape is not quite clear, but it now seems that the longer axis of the oriented lakes is perpendicular to the direction of the prevailing summer winds.

Thermokarst lakes also develop at the expense of ice wedges (see ICE WEDGE AND RELATED STRUCTURES). If a pool appears above an ice wedge (often as a result of a topographic change induced by the growth of the ice wedges), this pool warms the underlying soil and melts the top of the permafrost, i.e. the ice wedge. Because of this melting, the pool grows above the ice wedge, forming linear and polygonal troughs separating centres of the polygons (Plate 139). The centres, which contain less ice, are brought out into relief and evolve into conical mounds called 'thermokarst mounds'. Related to melting ice wedges, beaded streams may develop; these are characterized by narrow reaches linking pools or small lakes. The pools occur at the junction of the ice wedges.

In Siberia, well-known thermokarst forms are the ALASes, thermokarstic depressions with steep sides and flat bottoms covered by grassland, where shallow lakes often exist. Such depressions are generally round or oval, 3–40 m deep and 0.1–15 km long. They occupy 40–50 per cent of the surface of

Plate 139 Thermokarst in Siberia. Vegetation was cut near the road and a part of the active layer was stripped off, causing melting of a polygonal net of ice wedges and consequent formation of thermokarst mounds

the Lena and Aldan terraces in central Yakutia (Washburn 1979: 274) and they result from the melting of exceedingly ice-rich permafrost developed in the silty cover of the terraces.

Melting PINGOS induce formation of closed depressions surrounded by ramparts; such depressions are also thermokarstic. In contemporary permafrost zones, they are called 'collapsed pingos' or 'pingo remnants'. Generally, pingos begin to melt at their top because of the cracks relating to their growth.

Melting PALSAS also give birth to depressions that are visible only with regards to the areas raised by permafrost. When permafrost has disappeared, only very short-lived, shallow depressions remain marked by vegetation different from that on both previously unfrozen peatlands and the remnants of the peaty permafrost landforms.

Lithalsas (the same forms as palsas, but without any cover of peat, and developed in mineral soil), after melting leave, like pingos, depressions surrounded by ramparts. However, unlike pingos, but in the same way as palsas, lithalsas appear as numerous and almost adjacent forms.

On slopes, the most spectacular thermokarst forms are the 'retrogressive thaw slumps'. Described from the boreal forest as well to the High Arctic, they result from a process initiated by thawing of ground ice. It begins by the slide of the active layer on the permafrost table, which acts as a lubricated slip plane for movement and controls the depth of the failure plane. This process produces semicircular hollows opening downslope and usually less than 2 m high (French 1996: 119). Further thawing of permafrost produces steep slopes as much as 8 m high.

A peculiar thermokarstic phenomenon is fluviothermal erosion, namely erosion by the water of rivers, lakes or sea, attacking the permafrost not only by mechanical erosion processes, but also by its warmth which melts the ice. Such erosion is rapid and causes undercutting of riverbanks, particularly in sandy layers. It forms thermo-erosional niches at the floodwater level.

Thermokarstic phenomena have climatic or local causes. Climate warming, by increasing the thickness of the active layer, leads to the melting of the upper permafrost and to thermokarstic phenomena. However, response to climate warming is complex: besides temperature, snowfall also varies and changes in vegetation occur. The immediate response is not obvious. Nevertheless because of the global change, thermokarstic phenomena are finally to be feared.

But more often than not, permafrost is melting because of local causes. Thermokarstic phenomena, for instance, result from destruction or change of the vegetation cover as a consequence of forest fires or human action. Vegetation plays a complex role, generally protecting soil against warmth more than against cold. It acts in winter by trapping snow between branches and impeding insulation of the soil, in summer by protecting the soil by its shade, or by decreasing air circulation, by increasing evaporation, etc.

On areas weakened by excess ice in the soil, buildings, roads, airports, pipelines, etc. pose severe problems that engineers try to solve by impeding the melting of the permafrost. Houses and pipelines are built on piles; roads and airstrips are embanked above the ground in order to lift the permafrost surface; sometimes cooling devices are put into the ground to radiate its warmth to the surface.

Surveying fossil thermokarstic forms in regions where permafrost existed in the past has aroused interest in many scientists. Such forms already mentioned are pingo and lithalsa scars, recognizable by the ramparts surrounding the depressions (Pissart 2000: 344). Traces of thermokarstic collapses have disappeared because of the general subsidence induced by the melting of the whole permafrost. Only remnants of peculiar sediments deposited in those temporary hollows reveals their former existence. However thermokarstic explanations are often invoked without such observations. Their origin remains uncertain.

References

Czudek, T. and Demek, J. (1970) Thermokarst in Siberia and its influence on the development of lowland relief, *Quaternary Research* 1, 103–120.

French, H.M. (1996) *The Periglacial Environment*, Harlow: Longman.

Pissart, A. (2000) Remnants of Lithalsas of the Hautes Fagnes, Belgium: a summary of present-day knowledge, *Permafrost and Periglacial Processes* 11(4), 327–355.

Washburn, A.L. (1979) *Geocryology. A Survey of Periglacial Processes and Environments*, London: Edward Arnold.

Further reading

Ballantyne, C.K. and Harris, C. (1994) *The Periglaciation of Great Britain*, Cambridge: Cambridge University Press.

Harris, S.A., French, H.M., Heginbottom, J.A., Johnston, G.H., Ladanyi, B., *et al.* (1988) *Glossary of Permafrost and Related Ground-Ice Terms*, Ottawa: National Research Council Canada, Technical Memorandum no. 142.

ALBERT PISSART

THRESHOLD, GEOMORPHIC

A geomorphic threshold can be defined as the critical condition at which a landform abruptly changes. The change can be the result of an external variable, that exceeds the stability of a landform at an extrinsic threshold, or the change at an intrinsic threshold can be the result of a progressive change of the landform itself.

Extrinsic thresholds have been recognized in many fields. Perhaps the best known is the threshold velocity required to set in motion sediment particles of a given size. With a continuous increase in velocity, a threshold velocity is reached at which sediment movement commences, and with a progressive decrease in velocity, a threshold velocity is encountered at which sediment movement ceases. The best known thresholds in hydraulics are described by the Froude and Reynolds numbers, which define the conditions at which flow becomes supercritical or turbulent. Particularly notable are the changes of BEDFORM characteristics at threshold values of stream power. In these examples, an external variable changes progressively thereby triggering an abrupt change within the affected system at an extrinsic threshold. That is, the threshold exists within the system, but it will not be crossed and change will not occur without the influence of an external variable. The word, threshold, describes the critical range of conditions over which these transitions occur.

Thresholds can also be exceeded when the external variables remain relatively constant, yet a progressive change of the landform itself renders it unstable, and failure occurs at an intrinsic geomorphic threshold. An example is long-term progressive weathering, that reduces the strength of slope materials until eventually there is slope adjustment and MASS MOVEMENT. Another example of an intrinsic threshold is provided by a typical sequence of morphologic changes resulting in the collapse of sandstone-capped cliffs. Beneath a vertical cliff of sandstone is a gentler slope of weak shale. Through time, the basal shale slope is eroded, which produces a vertical shale cliff beneath the sandstone cap. At some critical height, the cliff collapses and the cycle begins again. The episodic retreat of this type of escarpment is the result of the change in cliff morphology under essentially constant climatic, base level and tectonic conditions. Similarly, a meander can increase in amplitude until a cutoff occurs under constant hydrologic conditions.

Field and experimental work supports the concept of geomorphic thresholds, which have been used to explain the distribution of discontinuous gullies in semi-arid valleys. Discontinuous gullies, short gullied reaches of valley floors, can be related to the slope of the valley-floor surface. For example, the beginning of GULLY erosion in these valleys tends to be localized on steeper reaches of the valley floor, which are the result of sediment storage. For a region of uniform geology, land use and climate, a critical intrinsic threshold of valley slope exists above which the valley floor is unstable and subject to incision during floods.

Similar relationships can be established for other alluvial deposits. For example, trenching of ALLUVIAL FANS is common, and the usual explanation for fan-head trenches is renewed uplift of the mountains or climatic fluctuations. However, as the fan grows through continual deposition, the fan-head steepens until it exceeds a threshold slope, when trenching occurs. Experimental (see EXPERIMENTAL GEOMORPHOLOGY) studies of alluvial-fan growth confirm that periods of trenching alternate with deposition at the fan-head. Therefore, the fan-head trenching can occur as a result of the oversteepening of the fan-head, and it is the result of the exceeding of an intrinsic geomorphic threshold.

A similar example is provided by damaging debris-flow events along the Wasatch Mountains in Utah. In 1993, a storm triggered debris flows from some canyons, but not all. It was suggested that debris basins be constructed at the mouths of the active canyons. However, further investigation revealed that the active canyons had been flushed of sediment and were not a threat, whereas the inactive canyons were storing sediment, which at some future time would produce damaging debris flows. The storage and flushing of sediment in these canyons is similar to the storage of sediment and its incision in semi-arid valleys and on alluvial fans as an intrinsic threshold is exceeded.

The identification of an intrinsic geomorphic threshold has significant practical applications.

If, as in the study of discontinuous gullies and alluvial-fan trenches, the critical slope is identified (intrinsic threshold), then unfailed, but sensitive valley floors and fan-heads can be identified. In this way, preventive conservation can be practised, thereby preventing erosion rather than attempting to control it after incision has occurred.

The concept of intrinsic geomorphic thresholds, which involves landform change without a change in external controls, challenges the well-established geomorphic thesis that relatively abrupt landform change is the result of some climatic, base level, or land-use change. Therefore, the significance of the intrinsic geomorphic threshold concept for geomorphologists is that it makes them aware that abrupt erosional and depositional changes can be inherent in the normal development of a landscape and that a change in an external variable is not always required for a geomorphic threshold to be exceeded and for a significant geomorphic event to result.

Further reading

Begin, Z.B. and Schumm, S.A. (1984) Gradational thresholds and landform singularity, *Quaternary Research* 21, 267–274.

Coates, D.R. and Vitek, J.D. (eds) (1980) *Thresholds in Geomorphology*, London: Allen and Unwin.

Patton, P.C. and Schumm, S.A. (1975) Gully erosion, northwestern Colorado: a threshold phenomenon, *Geology* 3, 88–90.

Phillips, J.D. (2001) The relative importance of intrinsic and extrinsic factors in pedodiversity, *Annals of the Association of the American Geographers* 91, 609–621.

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

——(1979) Geomorphic thresholds: the concept and its applications, *Transactions of the Institute of British Geographers* 4, 485–515.

Schumm, S.A., Harvey, M.D. and Watson, C.C. (1984) *Incised Channels: Morphology, Dynamics, and Control*, Littleton, CO: Water Resources Publications.

Westcott, W.A. (1993) Geomorphic thresholds and complex response of fluvial systems: some implications for sequence stratigraphy, *Bulletin American Association of Petroleum Geologists* 77, 1,208–1,218.

STANLEY A. SCHUMM

TIDAL CREEK

A creek is an inlet in a shoreline, a channel in a marsh, or another narrow, sheltered waterway. Creeks occur extensively on MUD FLATS AND MUDDY COASTS, on MANGROVE SWAMPS and on SALTMARSH surfaces (Eisma 1998).

Tidal creeks often have a high drainage density because of the large volumes of water that they drain. Saltmarsh creek densities may be 40 km/km^2 (Pethick 1984). The morphology of the creeks is also often distinctive. Although some may bear a superficial resemblance to dendritic river channel networks, flow along them is bi-directional (French and Stoddart 1992; Pestrong 1965). They have a tendency to taper upstream and flare downstream (Fagherazzi and Furbish 2001), and their discharge is determined by the tidal prism. In areas with a large tidal range or rapid seaward progradation, creek systems may be markedly linear in form. In areas with cohesive sediments creeks have steep edges, whereas in sandier areas they tend to be shallower and wider.

References

Eisma, D. (1998) *Intertidal Deposits: River Mouths, Tidal Flats, and Coastal Lagoons*, Boca Raton, FL: CRC Press.

Fagherazzi, S. and Furbish, D.J. (2001) On the shape and widening of salt marsh creeks, *Journal of Geophysical Research* 106, 991–1,003.

French, J.R. and Stoddart, D.R. (1992) Hydrodynamics of saltmarsh creek systems: implications for marsh morphological development and material exchange, *Earth Surface Processes and Landforms* 17, 235–252.

Pestrong, R. (1965) The development of drainage patterns on tidal marshes, *Stanford University Publications in Geological Science* 10, 1–87.

Pethick, J. (1984) *An Introduction to Coastal Geomorphology*, London: Arnold.

A.S. GOUDIE

TIDAL DELTA

Tidal deltas are large sand bodies formed within, or in the vicinity of, tidal inlets. The latter may be associated with barrier island chains (see BARRIER AND BARRIER ISLAND) and the entrances to coastal lagoons (see LAGOON, COASTAL) or estuaries (see ESTUARY). Flood-tidal deltas form landward of the inlet mouth, under the influence of flood-tidal currents. Ebb-tidal deltas occur seaward of the inlet, predominantly under the influence of ebb-tidal currents and wave action.

The major morphological features of flood-tidal deltas typically include (after Hayes 1980): a seaward-dipping flood ramp, up which landward sand movement occurs through the migration of sand waves under the action of flood currents; subtidal flood channels, which extend into the inlet and which dissect the partly intertidal landward portion of the delta (the 'ebb shield'); marginal ebb-aligned spits; and spillover lobes formed by the action of ebb currents over the lower parts of the ebb shield.

Ebb-tidal deltas are usually comprised of: an ebb channel, maintained by strong tidal currents; linear bars, formed through wave-current interactions along the margins of the ebb channel; a terminal lobe formed at the distal (seaward) end of the ebb channel, where the tidal current diminishes; sandsheets (or 'swash platforms') formed by wave action adjacent to the ebb channel characterized by migrating swash bars; and marginal channels dominated by flood-tidal currents.

Studies of inlet morphometry have shown that the morphology of tidal deltas is related to tidal prism (itself a function of both tidal range and inlet geometry), the configuration of the inlet and adjacent shoreline (including the offshore bathymetry), wave climate, and the rate of littoral sediment transport. In micro-tidal areas, flood deltas are often better developed than their ebb counterparts, owing to the dominance of landward, wave-driven, sediment transport. Ebb delta morphology is generally more variable than that of flood deltas, owing to the importance of regional and local contrasts in wave climate (Boothroyd 1985), and due to the tighter coupling of delta processes with wider coastal morphodynamic behaviour. Ebb delta volume increases with tidal prism, and decreases with inlet width/depth ratio and wave energy. Under conditions of low wave energy, ebb deltas are typically more elongated and extend further seaward. These controls are interactive so that, for example, wave energy can be modified by the presence of a headland, which also influences both the tidal prism and the intensity of tidal flows. For a sample of seventeen natural inlets in North Island, New Zealand, Hicks and Hume (1996) showed that over 80 per cent of the variation in ebb delta volumes could be successfully predicted by an empirical equation incorporating spring tidal prism and the angle between the ebb channel and the adjacent shoreline. Other factors accounting for some of the observed variation in delta volume include wave energy, sediment grain size (finer sands are less likely to be retained in the vicinity of strong tidal current jets and are thus associated with smaller deltas), and the supply of sediment through littoral drift.

FitzGerald *et al.* (2002) have drawn attention to marked contrasts in the occurrence and morphology of tidal deltas along the coast of New England, associated with a large degree of variability in tidal range, wave energy, sediment supply and inlet origin and geometry. Flood-tidal deltas in meso-tidal inlets tend to have a classic horseshoe shape and a significant intertidal area. Those in micro-tidal inlets tend to be predominantly subtidal and digitate or multi-lobate in form, owing to the limited ability of weak ebb currents to rework their deposits. Ebb-tidal deltas are best developed in moderately sized mixed energy environments. In wave-dominated inlets, they are either absent or small and entirely subtidal.

Although flood-tidal deltas act as long-term sediment sinks, ebb-tidal deltas are more dynamically coupled to the morphodynamic adjustment of the adjacent coast. Important processes include partial wave sheltering by the delta sand body; wave refraction around the delta, causing trapping and storage of beach sediments; and sediment recirculation within the delta (Oertel 1977). Ebb-tidal deltas interrupt the continuity of along-coast sediment movement, and bypass sand across their inlets in discrete pulses. Hicks *et al.* (1999) have shown that such inlet/ebb-delta processes can be an important source of interdecadal variability in beach behaviour. Years when the delta is accumulating sand are associated with erosion of beaches on the downdrift side of the inlet. Conversely, in years when the delta releases sand, the same beaches experience accretion associated with a migrating sand pulse.

Tidal deltas have historically been exploited as a sand resource (including mining to supply BEACH NOURISHMENT schemes). Questions are now being asked over the sustainability of this practice and its implications for beach stability. Furthermore, the correspondence between delta volumes and tidal prism means that their sand-trapping function is potentially sensitive to sea level rise. Inlets with extensive intertidal areas might experience a significant increase in tidal prism with a rise in sea-level, thus leading to increased sand storage. This may have adverse consequences for the stability of adjacent beaches, especially downdrift of the inlet.

References

Boothroyd, J.C. (1985) Tidal inlets and tidal deltas, in R.A. Davis (ed.) *Coastal Sedimentary Environments*, 2nd edition, 445–532, New York: Springer-Verlag.

FitzGerald, D.M., Buynevich, I.V., Davis, R.A. and Fenster, M.S. (2002) New England tidal inlets with special reference to riverine-associated inlet systems, *Geomorphology* 48, 179–208.
Hayes, M.O. (1980) General morphology and sediment patterns in tidal inlets, *Sedimentary Geology* 26, 139–156.
Hicks, D.M. and Hume, T.M. (1996) Morphology and size of ebb-tidal deltas at natural inlets on open sea and pocket-bay coasts, North Island, New Zealand, *Journal of Coastal Research* 12, 47–63.
Hicks, D.M., Hume, T.M., Swales, A. and Green, M.O. (1999) Magnitudes, spatial extent, time scales and causes of shoreline change adjacent to an ebb tidal delta, Katikati Inlet, New Zealand, *Journal of Coastal Research* 12, 220–240.
Oertel, G.F. (1977) Geomorphic cycles in ebb deltas and related patterns of shore erosion and acretion, *Journal of Sedimentary Petrology* 47, 1,121–1,131.

J.R. FRENCH

TOMBOLO

A sandbar, barrier or spit that joins an island with a mainland or another island, resulting from longshore drift or the migration of an offshore bar toward the coast. Tombolos are constructive features (though ultimately ephemeral due to wave erosion), occurring along shorelines of submergence that are protected from large waves and where islands are common. Sediment supply is predominantly derived from the islands, yet some may also come from erosion of the shoreline, fluvial materials, underwater reefs and offshore glacial deposits. Several types of tombolo exist including single, double, multiple, forked, parallel and complex tombolos, all of which are reflective of the coastal system (e.g. wave mechanisms) from which they are derived. For example, double tombolos (two ridges extending to shore), often form in areas with seasonal shifts in longshore drift. Tombolos can restrict flow between the sea and intertidal zone, forming a lagoon (see LAGOON, COASTAL), and altering the local ecology. An example of a tombolo is Chesil Beach, which extends northwestwards from the Isle of Portland to the coast of Dorset, south England.

Further reading

Schwartz, M.L. (1972) *Spits and Bars*, Benchmark Papers in Geology, Stroudsburg, PA: Dowden, Hutchinson and Ross.

SEE ALSO: bar, coastal; barrier and barrier island; coastal geomorphology

STEVE WARD

TOR

Linton (1955: 470) describes a tor as 'a solid rock outcrop as big as a house rising abruptly from the smooth and gentle slopes of a rounded summit or broadly convex ridge'. Tors are in fact large, free-standing, residual masses of rock (Plate 140). The word derives from the Old Welsh word *twr* or *twrr* meaning heap or pile. Tors are most common – and most well known – in granitic rocks (e.g. Haytor Rocks, Dartmoor, England), but also occur in coarse sandstones, schists, dacites and dolerites, among other lithologies. Although perhaps best known in south-west England (Devon and Cornwall), tors occur on all continents. In Africa, tors are often known as castle koppies (or kopjes). The rocks in which they occur vary considerably in age, but it is thought that most tors formed during the Tertiary or Pleistocene. Tors may occur in any position in the landscape, but are most common in summit and spur locations (Gerrard 1978; Ehlen 1991). They may occur as single, massive exposures (e.g. Middle Staple Tor, Dartmoor), or consist of groups of individual outcrops clustered together (e.g. Great Mis Tor, Dartmoor). The latter configuration is most common in summit positions. They may also appear to be piles of very large, loose boulders (core stones), but a rock core anchored in bedrock is usually present in the centres of such exposures.

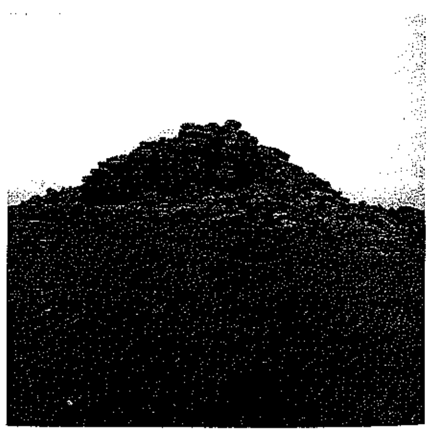


Plate 140 Middle Staple Tor, Dartmoor, south-west England. This tor is approximately 18 m long, 8 m tall and 10 m deep

Occasionally, the individual blocks in large summit tors form a pattern such that there is an elongate, open space, called an avenue, at the crest (e.g. Hound Tor, Dartmoor). Summit tors tend to be the largest, and those along valley sides tend to be the tallest (e.g. Vixen Tor, Dartmoor). Spur tors are most often the smallest. Sizes range from about one metre in height and a few metres in length and width to tens of metres in each dimension. Fallen rock debris, called clitter on Dartmoor, is often present at the bases of tors as well as for some distance downslope. Much of the clitter has been moved downslope by periglacial processes (e.g. stone lines on the west side of Middle Staple Tor, Dartmoor).

It is generally accepted that tor shapes and locations in all lithologies are controlled by JOINTING. Each tor typically contains three major joint sets (and two to five minor ones), one horizontal or gently dipping set defining the top, and two vertical or steeply dipping sets defining the sides of the tor. Diagonal joints may also be present (e.g. Great Tor, Dartmoor), but they are not common. Tors have a variety of appearances depending upon the arrangement of the joints that form them. They may be quite massive blocks of rock, either very rounded or blocky in appearance (Plate 140). They can be lamellar in form (Haytor Rocks, Dartmoor). They can also be tall and narrow, almost pinnacles (e.g. Bowerman's Nose, Dartmoor). These variations in appearance are caused by changes in the distributions of the different types of joints in the tor. If vertical and horizontal joints occur with about equal frequency, the tor will be composed of joint blocks of approximately equal size and will appear massive and blocky. If horizontal or gently dipping joints occur in significantly greater numbers than vertical or steeply dipping joints, the tor will be lamellar in form. If horizontal or gently dipping joints are rare, the tor will be tall and narrow. Logan stones, joint-bounded, precariously balanced boulders that move when touched, can be found in association with all types of tors, but are most common in shorter tors where horizontal joints are closer together than vertical joints, producing an elongate, rectangular block.

Origin of tors

Various theories have been proposed for the origin of tors. One theory suggests that they are the product of atmospheric weathering – that wind, rain, freeze – thaw, salt crystallization and insolation round the shapes of exposed, angular rock outcrops produce rounded, bouldery tors (e.g. Palmer and Neilson 1962). This theory is generally discounted except in certain unusual and specific cases (e.g. Selby 1972).

A second theory for the origin of tors assumes they are the product of a two-stage process (the two-stage theory). In this theory, based on Linton's (1955) work on Dartmoor in south-west England, tors form in the subsurface by chemical weathering along joints, and are subsequently exposed by erosional stripping. Once exposed, they retain their rounded shape. Linton noted a buried tor in a small quarry near Two Bridges in central Dartmoor as unrefutable evidence of this theory. Others (e.g. Thomas 1974) have expanded Linton's theory to include multiple phases of weathering and denudation. The theory suggests that weathering progresses most rapidly where the distance between joints is narrow, and more slowly where it is wide. Once the weathered mantle, which is thickest where the joints were most closely spaced, is removed by erosion, a rock outcrop with relatively widely spaced joints remains. Ehlen *et al.* (1997) showed that this was in fact the case in the Granite Mountains, Wyoming, and Ehlen and Wohl (2002) provided additional evidence favouring this theory in their study of BEDROCK CHANNELS in the Colorado Front Range. The weathered mantle formed from granitic rocks is called growan, GRUS, or saprolite. This theory has received significant support from many workers over the years, such as Eden and Green (1971) in their work on Dartmoor and C.R. Twidale in his many papers primarily on granite landforms in Australia. The genetic nature of the two-stage theory, however, has made it difficult for many to accept.

The third theory (the scarp retreat theory), proposed by King (1949) and based on his work in southern Africa, is that tors are the product of lateral planation and pediment formation. In a later publication, King (1958) accepts that sub-skyline tors could result from chemical weathering and exhumation, and states that his 1949 theory refers only to skyline (i.e. summit) tors. Ollier and Tuddenham (1961) in their work on Australian INSELBERGS and Ojany's (1969) on Kenyan inselbergs, among others, have provided support for King's scarp retreat theory.

The final theory of tor formation (the periglacial theory) is that proposed by Palmer and

Radley (1961). Their studies of sandstone tors in north-east England, where there is no evidence of deep chemical weathering, suggest that the tors were isolated from free faces along joint planes, and then rounded and fretted by subsequent atmospheric denudation. Palmer and Nielson (1962) studied the Dartmoor tors and suggested that – because core stones are lacking, there is no evidence of a DEEP WEATHERING PROFILE, and atmospheric weathering can account for the rounded shapes – the Dartmoor tors have a periglacial origin. They proposed a three-stage theory of periglacial tor formation. The application of this theory to the Dartmoor tors, however, is generally discounted.

The favoured theories among those described above are the modified two-stage theory first presented by Linton and King's scarp retreat theory, and much work has been done since they first published their work to support one or the other theory, as noted above. But perhaps the most reasonable approach to the origin of tors is that tors form by different processes in different environments (i.e. the principle of equifinality) as suggested by Brunsden (1964) and Thomas (1974), among others.

References

Brunsden, D. (1964) The origin of decomposed granite on Dartmoor, in I.G Simmons (ed.) *Dartmoor Essays*, Exeter: Devonshire Association for the Advancement of Science, Literature and Art, 97–116.

Eden, M.J. and Green, C.P. (1971) Some aspects of granite weathering and tor formation on Dartmoor, England, *Geografiska Annaler* 53, 92–99.

Ehlen, J. (1991) Significant geomorphic and petrographic relations with joint spacing in the Dartmoor Granite, southwest England, *Zeitschrift für Geomorphologie* 35, 425–438.

Ehlen J. and Wohl, E. (2002) Joints and landform evolution in bedrock canyons, *Transactions, Japanese Geomorphological Union* 23, 237–255.

Ehlen, J., Gerrard, J. and Zen, E. (1997) Joint spacing and landform evolution: the Granite Mountains, WY, *Geological Society of America Abstracts with Program* 29, A36.

Gerrard, A.J.W. (1978) Tors and granite landforms of Dartmoor and eastern Bodmin Moor, *Proceedings of the Ussher Society* 4, 201–210.

King, L.C. (1949) A theory of bornhardts, *Geographical Journal* 112, 83–87.

——(1958) Correspondence on the problem of tors, *Geographical Journal* 124, 289–291.

Linton, D. (1955) The problem of tors, *Geographical Journal* 121, 470–487.

Ojany, F. (1969) The inselbergs of eastern Kenya with special reference to the Ukambani area, *Zeitschrift für Geomorphologie* 13, 196–206.

Ollier, C.D. and Tuddenham, W.G. (1961) Inselbergs of central Australia, *Zeitschrift für Geomorphologie* 5, 257–276.

Palmer, J. and Neilson, R.A. (1962) The origin of granite tors on Dartmoor, Devonshire, *Proceedings of the Yorkshire Geological Society* 33, 315–340.

Palmer, J. and Radley, J. (1961) Gritstone tors of the English Pennines, *Zeitschrift für Geomorphologie* 5, 37–52.

Selby, M.J. (1972) Antarctic tors, *Zeitschrift für Geomorphologie* 13, 73–86.

Thomas, M.F. (1974) Granite landforms: a review of some recurrent problems in interpretation, in E.H. Brown and R.S. Waters (eds) *Progress in Geomorphology*, Institute of British Geographers Special Publication No. 7, 13–35.

Further reading

Gerrard, A.J. (1988) *Rocks and Landforms*, London: Unwin Hyman.

Twidale, C.R. (1982) *Granite Landforms*, Amsterdam: Elsevier.

SEE ALSO: exhumed landform; granite geomorphology; inselberg; rock control; salt weathering; spheroidal weathering; weathering

JUDY EHLEN

TOREVA BLOCK

Large masses of relatively stratigraphically coherent rock that have slipped down a cliff or mountain side upon normal listric faults, and has rotated backwards toward the parent cliff (Reiche 1937). The blocks can measure beyond 600 m in thickness and lateral extent, and are end members for this landslide type. Some blocks lie close to the parent cliff whereas others may have slipped several hundred kilometres from their sources. Their emplacement age remains uncertain, though probably in different (humid) climates in the Pleistocene.

Reference

Reiche, P. (1937) The toreva-block, a distinctive landslide type, *Journal of Geology* 45, 538–548.

SEE ALSO: mass movement

STEVE WARD

TRACER

Tracer techniques enable geomorphologists to quantify the movement of Earth materials (whole systems, individual particles, water) and provide

data that enable them to model the movement of these materials through a range of Earth systems. Applications have focused on four major research areas.

1. Studies of whole system behaviour (e.g. measurement of soil creep, mass movement and mudflows on hillslopes; measuring rates of movement and internal deformation of glaciers, measuring the surface deformation of volcanoes) (Meier 1960; Carson and Kirkby 1972; Anderson and Finlayson 1975; Goudie 1990).
2. Studies of coarse sediment transport on hillslopes, in fluvial systems and in littoral zones (e.g. determination of entrainment thresholds, transport distances, particle size and shape controls on sediment transport) (Sear *et al.* 2000).
3. Studies of fine particle transport and sediment provenance (e.g. airborne particles, hillslope erosion and soil redistribution, quantifying sediment-associated contaminant movement and determining the provenance of a range of dated fluvial, limnic, aeolian and marine deposits) (Foster and Lees 2000).
4. Tracing water movement and flow pathways (e.g. through soil profiles, groundwater and cave systems and estimating flood wave travel times in rivers) (Leibundgut 1995; Kranjc 1997).

While by no means exhaustive, this brief list of examples demonstrates the remarkable breadth of tracer applications in geomorphology.

Whole system tracer studies have made use of inert or active surface and subsurface markers (e.g. three dimensional Global Positioning Systems; steel/aluminium rods or spheres; distinctly painted stones or pebbles) whose movement can be monitored directly by satellite tracking, by field re-survey or by using remotely sensed images (e.g. aerial photographs).

Studies of coarse sediment transport may use 'passive' natural materials that are marked in some way to allow identification, recovery and measurement (e.g. painted pebbles, pebbles drilled with bar magnets or impregnated/coated with radionuclides). Alternatively, a range of natural or artificial materials may contain radio transmitters to allow rapid location even if the particle is buried.

Studies of fine particle movement may use 'exotic' materials that have characteristic signatures (e.g. fine-grained magnetic powders; fluorescent micro-spheres) that can be detected by field survey or by field sampling and laboratory analysis. Natural properties of environmental materials can also be used if their signatures (e.g. geochemical, mineralogical, mineral magnetic, stable isotope, radionuclide) characterize a distinct source of origin (e.g. geological or lithological units, soil types, topsoils/subsoils).

Tracing water movement and flow pathways has made use of fine-grained neutrally buoyant materials that move in the same way as the liquid (e.g. *Lycopodium* spores) or a wide range of (fluorescent) soluble dyes (e.g. Rhodamine) that can be added to cave or groundwater systems or to the surface of a soil profile and which may be detected at low concentrations.

Whichever method is appropriate to the research problem, a number of assumptions underpin the use of all tracers whatever the field of study. Before looking at an example, we need to ask:

What makes a good tracer?

There are a number of factors which need to be taken into account in order to ensure a successful outcome in tracer experiments:

THE TRACER DOES NOT INTERFERE WITH OR ALTER THE PROCESS BEING MEASURED

In many cases this is difficult to achieve. For example, digging soil pits and excavating or drilling holes in glaciers, installing metal pins to estimate rates of soil creep or ice deformation, or installing piezometer tubes in a mudflow, directly disturb the immediately surrounding environment. While widely used, these and other intrusive methods of measurement are subject to unknown and largely unknowable errors.

THE TRACER IS REPRESENTATIVE OF THE SYSTEM BEING MEASURED

Artificial coarse and fine tracers may differ in size, shape and density from the natural materials being investigated. All these factors are important. For coarse particle studies, for example, entrainment and settling velocities are functions of the mass, density and shape of the particle. In fine particle studies (especially in the silt and clay-sized fractions), particle interactions are important and many natural materials move as aggregates rather than as individual particles. Replacement of natural by artificial materials may poorly replicate particle interactions.

THE TRACER CAN BE RECOVERED AND IDENTIFIED

Historically, poor tracer recovery in coarse sediment studies has posed a major problem, especially if particles have become buried. Poor recovery leads to problems in interpreting results in a statistically meaningful way. Tracers marked with surface paints, for example, are often difficult to relocate even if only buried to a few centimetres depth. Placement of pebbles in high-energy environments can lead to abrasion and loss of paint cover, while brightly coloured pebbles on a beach undoubtedly attract the attentions of young children often leading to redistribution in a manner no natural geomorphological process could explain. More recent developments in tracer technologies have used magnetic or radio-tracers emplaced in locally derived material so that even buried particles have a better chance of being found and do not attract unwarranted attention. It is essential that, whatever the marking system used, the tracer will last the lifetime of the project, that the identity is resistant to removal and that each individual particle has an unequivocal identity (Sear *et al.* 2000).

THE TRACER IS TRANSPORTED AND DEPOSITED IN THE SAME WAY AS THE MATERIAL BEING STUDIED

In many situations, coarse particles or other large material, can be identified by marking the surface with paint on which a code is added for later identification. Painted surfaces, however, may change the surface roughness and porosity characteristics of the original material leading to changes in buoyancy over relatively short timescales. Alternatives include the use of exotic (non-local) materials in order to aid identification and recovery but again differences in shape, density, porosity, buoyancy and/or surface roughness may lead to errors in experimental results. While magnetically or radio-tagged natural or 'exotic' particles have improved recovery rates, tracer physical characteristics may not exactly match those of the natural materials whose behaviour they are manufactured to represent. Fine sediment studies that use natural tracer characteristics may also prove problematic since erosion and sediment transport processes are particle-size selective. In consequence, concentrations of many natural tracers (e.g. heavy metals, nutrients, radionuclides, mineral magnetic signatures) increase with a decrease in particle size and an increase in particle specific surface area and for which a correction factor is often required. Additional considerations must be given to changing environmental conditions during transport. This is especially true in aquatic systems where changes in pH, redox potential (Eh) and salinity may drive adsorption and/or desorption reactions during sediment transport (Horowitz 1991).

LONG-STERM STORAGE

Once deposited, fine-grained aeolian, marine, fluvial and limnic sediments can undergo a range of complex transformations. Interruptions in loess or floodplain deposition, for example, often leads to periods of soil development while changes in the trophic status of lakes and estuaries may again lead to changes in pH and Eh driving the release of many natural tracers into the water column. It cannot therefore be assumed that natural tracer properties remain unchanged with time (e.g. Foster *et al.* 1998)

An example of fine sediment tracing for interpreting erosion processes

Erosion processes operating in a grazed paddock in New South Wales, Australia were analysed by Wallbrink and Murray (1993) using a rainfall simulator and two radionuclides that adsorb strongly to sediment and label different parts of the soil profile in different ways. Figure 168a and b show the vertical distribution of ^{137}Cs and ^{7}Be in a typical soil profile. ^{137}Cs has a thirty-year half-life and has been detected in the environment since atmospheric testing of thermonuclear weapons in the early 1950s (Higgitt 1995). Following the international treaty banning atmospheric nuclear weapons testing in 1963, the southern hemisphere has received little ^{137}Cs fallout. ^{7}Be is continuously produced in the upper atmosphere by cosmic ray bombardment but has a short half-life (53 days) in comparison with ^{137}Cs. Figure 168c shows how the ^{7}Be and ^{137}Cs activities of sediment produced by four different erosion processes (sheet, gully floor, gully collapse, rill erosion) could produce suspended sediment with different combinations of the two signatures. Gully collapse would result in low activities of both nuclides since the gully wall does not receive ^{7}Be fallout and the depth penetration of ^{137}Cs in the soil profile is less than 10 cm. The majority of the gully wall sediments are not labelled with either radionuclide. By contrast, sheet erosion would produce sediment with high activities of both radionuclides. Gully floor

sediments would be labelled with ^{7}Be (since it is produced continuously) but would have no ^{137}Cs (since in this case the gully developed after 1963). Rill erosion would produce sediment strongly labelled with ^{137}Cs, but the shallow depth penetration of ^{7}Be in the soil profile would lead to a dilution of ^{7}Be activities as it mixes with sediment derived from deeper in the soil profile. By collecting suspended sediment during rainfall simulation experiments generating surface runoff, Figure 168d shows that the model of high ^{7}Be and ^{137}Cs activity is supported by the experimental results.

References

Anderson, M.G. and Finlayson, B.L. (1975) *Instruments for Measuring Soil Creep*, BGRG Technical Bulletin No.16, Norwich: Geo Abstracts.

Carson, M.A. and Kirkby, M.J. (1972) *Hillslope Form and Process*, Cambridge: Cambridge University Press.

Foster, I.D.L. and Lees, J.A. (2000) Tracers in geomorphology: theory and applications in tracing fine particulate sediments, in I.D.L. Foster (ed.) *Tracers in Geomorphology*, 3–20, Chichester, Wiley.

Foster, I.D.L., Lees, D.E., Owens, P.N. and Walling, D.E. (1998) Mineral magnetic characterisation of sediment sources from an analysis of lake and floodplain sediments in the catchments of the Old Mill Reservoir

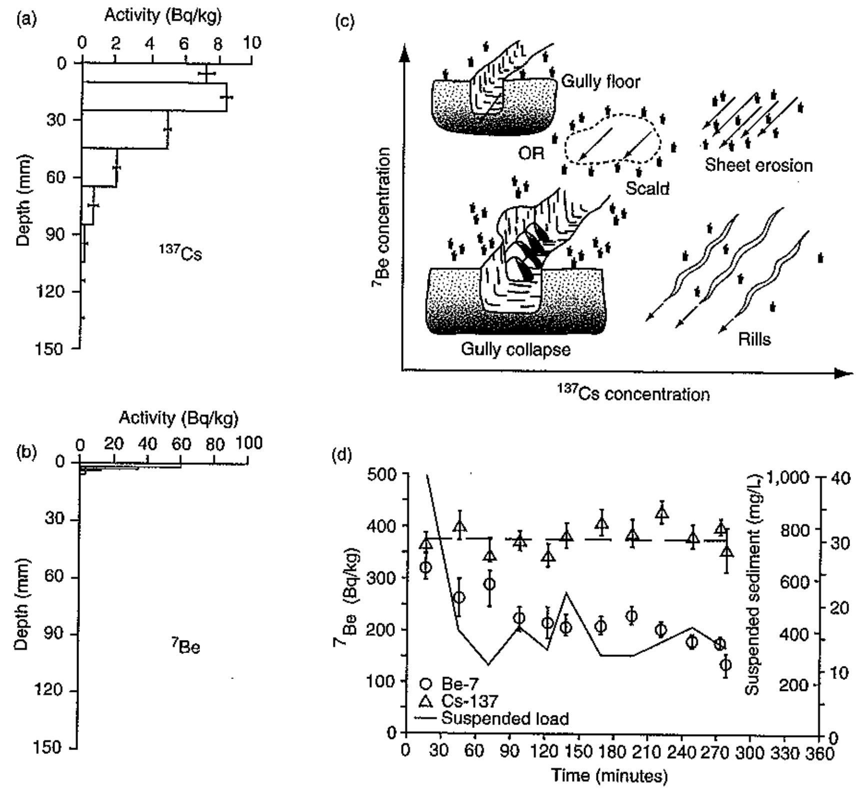


Figure 168 The distribution of (a) ^{137}Cs, and (b) ^{7}Be in the soil profile; (c) a model of how sediment would be labelled with ^{137}Cs and ^{7}Be depending on the erosion process involved; and (d) the results of a rainfall simulation experiment generating surface erosion

Source: Wallbrink and Murray (1993). © John Wiley and Sons Limited. Reproduced with permission.

and Slapton Ley, South Devon, UK, *Earth Surface Processes and Landforms* 23, 658–703.

Goudie, A. (ed.) (1990) *Geomorphological Techniques*, London: Allen and Unwin.

Higgitt, D.L. (1995) The development and application of Cs-137 measurements in erosion investigations, in I.D.L. Foster, A.M. Gurnell and B.W. Webb (eds) *Sediment and Water Quality in River Catchments*, 287–305, Chichester: Wiley.

Horowitz, A. (1991) *A Primer on Sediment Trace Chemistry*, Chelsea, MI: Lewis.

Kranjc, A. (1997) *Tracer Hydrology 97*, Rotterdam: Balkema.

Leibundgut, Ch. (1995) *Tracer Techniques for Hydrological Systems*, IAHS Publication 229, Wallingford: International Association of Hydrological Sciences Press.

Meier, M.F. (1960) Mode of flow of Saskatchewan Glacier, *US Geological Survey Professional Paper* No 351, Washington, DC.

Sear, D.A., Lee, M.W., Oakey, R.J., Carling, P.A. and Collins, M.B. (2000) Coarse sediment tracing technology in littoral and fluvial environments, in I.D.L. Foster (ed.) *Tracers in Geomorphology*, 21–55, Chichester: Wiley.

Wallbrink, P.J. and Murray, A.S. (1993) Use of fallout radionuclides as indicators of erosion processes, *Hydrological Processes* 7, 297–304.

Further reading

Peters, N.E., Hoehn, E., Leibundgut, Ch., Tase, N. and Walling, D.E. (eds) (1993) *Tracers in Hydrology*, IAHS Publication 224, Wallingford: IAHS Press.

Rukin, N., Hitchcock, M., Streetly, M., Al Faihani, M. and Kotoub, S. (1994) The use of fluorescent dyes as tracers in a study of artificial recharge in Qatar, in E.M. Adar and Ch. Leibundgut (eds) *Application of Tracers in Arid Zone Hydrology*, IAHS Publication 232, 67–78, Wallingford: IAHS Press.

Verusob, K.L., Fine, M.J. and TenPas, J. (1993) Pedogenesis and palaeoclimate interpretation of the magnetic susceptibility of the Chinese loess-palaeosol sequences, *Geology* 21, 1,011–1,014.

IAN D.L. FOSTER

TRANSGRESSION

A transgression is the movement of mean sea level (MSL) in an upward direction, while any lowering of MSL is called a regression. A movement of MSL, as the datum for wave and tidal activity, generates the potential for change in areas above and below the datum. Changes in the position of MSL are usually relative. A transgression is not solely due to MSL moving, as it is possible that the land surface is also moving relative to MSL. Long-term MSL changes can be due to both eustatic (changes in volume of sea water in the oceans) and isostatic (vertical movement of land) causes. A transgression is specified when the net balance between these two processes resolves in a relative rise in mean sea level (RSLR). Eustatic changes generally relate to climatic changes and their control on (1) unit ocean volume contraction/expansion due to atmospheric-ocean heat exchange (the steric effect); (2) evaporation rates of sea water; and (3) variations in the hydrological cycle by which free water is either locked up as terrestrial ice, or terrestrial ice melting to release water back into the oceans.

Quaternary transgressions usually relate to warming phases by which glacier ice melts and sea surface temperatures rise, e.g. the early to mid-Holocene in which atmospheric warming contributed to a major transgressive phase occurring in the mid-latitudes (Pirazzoli 1996). Although the Late Glacial was marked by rapid positive eustatic change, it was also a period of rapid upward crustal lift in the mid-upper latitudes, due to the release of terrestrial ice pressure through ice melting. The near-exponential early rapid isostatic rise in land level meant that eustatic change was more than exceeded, in effect inducing a regression or relative fall in MSL. This was only reversed as the isostatic rate diminished and the eustatic effect maximized some time in the mid-Holocene, to induce a well-recognized transgression (e.g. the Flandrian in north-west Europe). The British Isles can be crudely characterized as showing two main zones of response in the mid-Holocene; a northern zone in which the transgression peaked above present-day MSL (*c*.5–6 ka BP) and then switched to a regressive phase; and a southern zone showing a continuing transgressive phase to the present day but associated with a decelerating RSLR rate since the mid-Holocene. Local isostatic effects tend to modulate a regional eustatic signal and complicate the general picture at any one place. Contemporary concerns with the effects of accelerating global climate change are reflected in the forecasts for rising MSL over the next century, with RSLR up to five times the current rates being forecast. This will result in a major new transgressive phase, regardless of current crustal changes, as these modern eustatic changes are linked to global climate change, while the endogenetic changes required to generate crustal isostatic responses are unaffected.

Transgressions are a common element of geological-scale change. In recent decades, litho-stratigraphies have been interpreted via sequence stratigraphy by which rising and falling sea levels have been used to link subaerial erosion sources to submarine deposition sinks. The timescale of such deposition relates to transgressions ($>10^6$ years duration) as a consequence of: (1) slow orogeny and oceanic basin volume reduction; (2) massive sediment transfers due to landscape denudation causing isostatic readjustment; and (3) probable climate change. The speed of the Holocene transgression (millennia-scale) stands in contrast to these earlier episodes, and emphasizes the idiosyncratic conditions associated with Quaternary deglaciation-induced transgressions.

A transgression leads to inundation of the coastal zone, the onshore extent of which depends on the overall coastal slope angle. The sedimentary expression of the transgression is dependent on both the non-rectilinearity of the coastal slope setting the template for deposition, and availability of free sediment. The penetration distance of the transgression is not solely that of inundation, as the transgression carries wave and tidal activity that reworks the sediment mantle beyond the initial flooding limit. Bruun (1962) has attempted to specify the predictive relationship between RSLR and shoreline retreat, through what is now controversially termed the BRUUN RULE.

The concept of the 'Erosion Front' defines the spatial variation of coastal activity experienced from the quiescent leading edge of the front that moves up estuary, through the high energy breaking wave zone, and the trailing edge that has tidal reworking of near shore and beach face (Carter *et al.* 1992). The basal extent of the transgression is identified by an erosional surface also known as a ravinement. Much interest is centred on the way in which a transgression aids sediment-reworking into distinctive coastal morphologies, e.g. the development of both sand and gravel barriers and associated back-barrier intertidal sediment stores and marshes. Some dune investigators believe that transgressive conditions were requisite for the major coastal dunes associated with the Holocene, though a counter argument exists in which dunes are also associated with regressions. Local excess deposition can generate an apparent regressive signature by outweighing transgressive tendencies, though to achieve this sediment has to be derived from elsewhere alongshore leading to overall shoreline retreat. Key issues still to be resolved relate to transgressive migration rate and coastal morphology stability. Can saltmarsh growth match fast RSLR? Can barriers rollover and maintain longshore continuity under fast RSLR (Jennings *et al.* 1998)? These are critical issues for coastal communities dependent on maintaining sustainable natural coastal morphologies as coastal defences in the face of future extreme rates of RSLR.

References

Bruun, P. (1962) Sea-level rise as a cause of shore erosion, *Journal of Waterways, Harbour Division*, American Society of Civil Engineers 88 (WW1), 117–130.

Carter, R.W.G., Orford, J.D., Jennings, S.C., Shaw, J. and Smith, J.P. (1992) Recent evolution of a paraglacial estuary under conditions of rapid sea-level rise: Chezzetcook Inlet, Nova Scotia, *Proceedings of the Geological Association* 103, 167–185.

Jennings, S.C., Orford, J.D., Canti, M., Devoy, R.J.N. and Straker, V. (1998) The role of relative sea-level rise and changing sediment supply on Holocene gravel barrier development; the example of Porlock, Somerset, UK, *Holocene* 8, 165–181.

Pirazzoli, P. (1996) *Sea Level Changes: The Last 20000 Years*, Chichester: Wiley.

JULIAN ORFORD

TREE FALL

The falling over of trees during high winds is a significant factor in the translocation of material, churning of soils (Schaetzl 1986) disruption of strata, and the development of mound and pit micro-topography (Denny and Goodlett 1956). Deep-rooting trees affect topography to a greater degree than shallow-rooting trees, and when they are blown over leave deep pits and high mounds (Veneman *et al.* 1984). In mixed hardwood forests in Ithaca, New York, USA, mounds were typically 0.48–0.60 m high and pits 0.20–0.41 m deep (Beatty and Stone 1986). Windthrow appears to be a more prevalent phenomenon on deeper soils where there is a sharp contrast between the fine soil material and the underlying stony horizons (Boyd and Webb 1981). Tree fall is also a major contributor to the development of log steps and LARGE WOODY DEBRIS in forest streams (Marston 1982).

References

Beatty, S.W. and Stone, E.L. (1986) The variety of soil microsites created by tree falls, *Canadian Journal of Forest Research* 16, 539–548.

Boyd, D.M. and Webb, T.H. (1981) The influence of the soil factor on tree stability in Balmoral Forest, Canterbury, during the gale of August 1975, *New Zealand Journal of Forestry* 26, 96–102.
Denny, C.S. and Goodlett, J.C. (1956) Microrelief resulting from fallen trees, *US Geological Survey Professional Paper* 288, 59–66.
Marston, R.A. (1982) The geomorphic significance of log steps in forest streams, *Annals of the Association of American Geographers* 72, 99–108.
Schaetzl, R.J. (1986) Complete soil profile inversion by tree uprooting, *Physical Geography* 7, 181–189.
Veneman, P.L.M., Jacke, P.V. and Bodine, S.M. (1984) Soil formation as affected by pit and mound microrelief in Massachusetts, USA, *Geoderma* 33, 89–99.

A.S. GOUDIE

TRIMLINE, GLACIAL

Reconstruction of the geometry of former ice sheets and ice caps from geomorphological evidence is based on the identification of trimlines, defined by the maximum level to which GLACIERS or ICE SHEETS have eroded or 'trimmed' bedrock or debris in a valley hillslope (Ballantyne and Harris 1994). The sharpness of this boundary depends on the effectiveness of GLACIAL EROSION, the degree of frost WEATHERING after its formation, and the downslope MASS MOVEMENT during and after DEGLACIATION. The formation of weathering boundaries and glacial trimlines are open to four possible hypotheses (Figure 169):

1 Summit blockfields (see BLOCKFIELD AND BLOCKSTREAM) are formed by *in situ* rock weathering over a longer timescale than the Holocene, representing a glacial trimline marking the maximum altitude to which glacial erosion has eroded or 'trimmed' a pre-existing cover of REGOLITH/frost-shattered debris.
2 Frost weathering on high ground may reflect more pronounced breakup of rock at high altitudes under periglacial (see PERIGLACIAL GEOMORPHOLOGY) conditions, particularly during downwasting of an ice sheet and subsequent valley glaciation.
3 The initial stage of downwastage of ice sheets/ice caps may be accompanied by frost (see FROST AND FROST WEATHERING) shattering of exposed rock which ceases when the climate warms. The limit between frost-weathered and glacially abraded terrain represents the upper limit of glacier ice at the time of this thermal transition.
4 The weathering limit may represent a thermal boundary within a former ice sheet or ice cap, with *in situ* frost-weathered debris surviving under a cover of cold-based (basal temperature below the PRESSURE MELTING POINT) ice on high ground, whereas lower areas experience scouring by warm-based (basal temperature at the pressure melting point) glaciers.

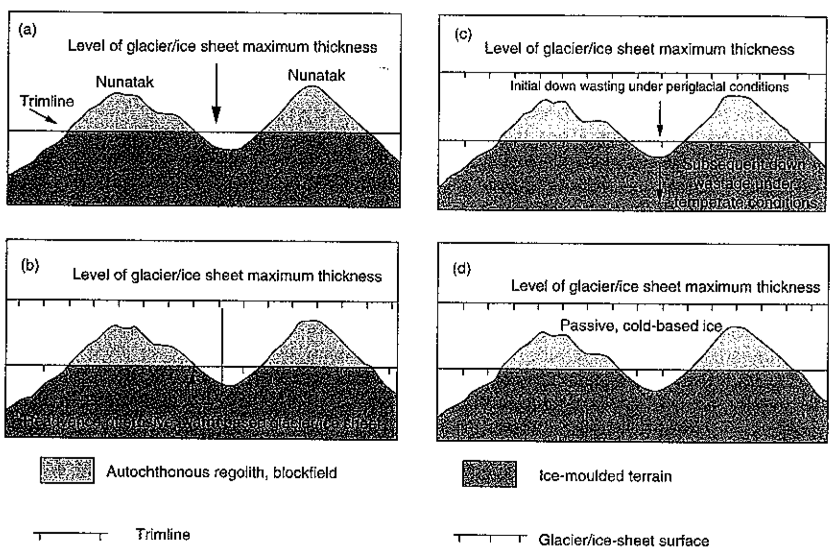


Figure 169 Four hypotheses of formation of glacial trimlines: (a) a glacial trimline representing the upper surface of an ice sheet at its maximum thickness; (b) a trimline cut by a glacial readvance during overall ice-sheet downwastage; (c) trimline formed during an initial period of ice-sheet downwastage under periglacial conditions; (d) weathering limit representing a thermal boundary between cold-based (temperature below pressure melting point) ice and a warm-based (temperature at the pressure melting point) ice within a former ice sheet (modified from Ballantyne *et al.* 1998)

Various approaches have been adopted to test these hypotheses. These involve analyses of weathering characteristics of bedrock and soils above and below the weathering limit/trimlines, and reconstruction of their altitudinal trend. Techniques employed are SCHMIDT HAMMER measurements of rock surface hardness, measurements of surface ROUGHNESS, measurements of differential relief of adjacent minerals, depth of open horizontal dilation (stress-release) joints using a graduated probe, studies of clay mineral assemblages and mineral magnetic signatures, and COSMOGENIC DATING.

References

Ballantyne, C.K. and Harris, C. (1994) *The Periglaciation of Great Britain*, Cambridge: Cambridge University Press.
Ballantyne, C.K., McCarroll, D., Nesje, A., Dahl, S.O. and Stone, J. (1998) The last ice sheet in north-west Scotland: reconstruction and implication, *Quaternary Science Reviews* 17, 1,149–1,184.

Further reading

Brook, E.J., Nesje, A., Lehman, S.J., Raisbeck, G.M. and Yiou, F. (1996) Cosmogenic nuclide exposure ages along a vertical transect in western Norway: implication for the height of the Fennoscandian ice sheet, *Geology* 24, 207–210.
Grant, D.R. (1977) Altitudinal weathering zones and glacial limits in Western Newfoundland, with particular reference to Gros Morne National Park, *Geological Survey of Canada Paper* 77–1A, 455–463.
Locke, W.W. (1995) Modelling the icecap glaciation of the northern Rocky mountains of Montana, *Geomorphology* 14, 123–130.
Nesje, A., Dahl, S.O., Anda, E. and Rye, N. (1988) Blockfields in southern Norway; significance for the Late Weichselian ice sheet, *Norsk Geologisk Tidsskrift* 68, 149–169.
Stone, J.O., Ballantyne, C.K. and Fifield, L.K. (1998) Exposure dating and validation of periglacial weathering limits, northwest Scotland, *Geology* 26, 587–590.

ATLE NESJE

TROPICAL GEOMORPHOLOGY

Tropical geomorphology has usually addressed conditions in the tropical forests and savannas, and although many areas within the tropics (23.5°N and S lat.) fall within the arid and semi-arid climates, these environments will not be included here. However, near-tropical climates extend to at least 30° lat., along the humid east coasts of Africa, Asia and the Americas and these areas have much in common with the tropics *sensu stricto* (Figure 170).

The study of geomorphology in the tropics cannot be separated from the wider history of the subject. During the early twentieth century most reports on tropical landscapes arose from geological investigations on behalf of former colonial powers: Britain, France, the Netherlands. Many observations concerned the extent and products of rock weathering (Falconer 1911; Scrivenor 1931). Reports also commented on unusual landscapes: extensive plains and great waterfalls in east and south Africa; high granite domes (or *inselberge*) in east and west Africa; the tower karst of South-East Asia. In addition, the widespread occurrence of *laterite*, long known from Buchanan's early work in India, attracted a lot of comment. Because the authors came from Europe, the exotic nature of tropical landforms was frequently emphasized.

In Davis's (1899) paper on 'The cycle of erosion' the evolution of landscapes was seen from the perspective of a humid temperate 'normality' and although Davis later addressed contrasts with semi-arid regions in western USA, the tropics were largely ignored. For fifty years the Davisian view was dominant in Britain, France and the USA and tropical landforms were seen as 'climatic accidents' (Cotton 1942) and peripheral to mainstream interests. In Germany, tropical landscapes were considered within a 'climatic geomorphology' (Thorbecke 1927; Büdel 1948), which led to specific hypotheses for landform

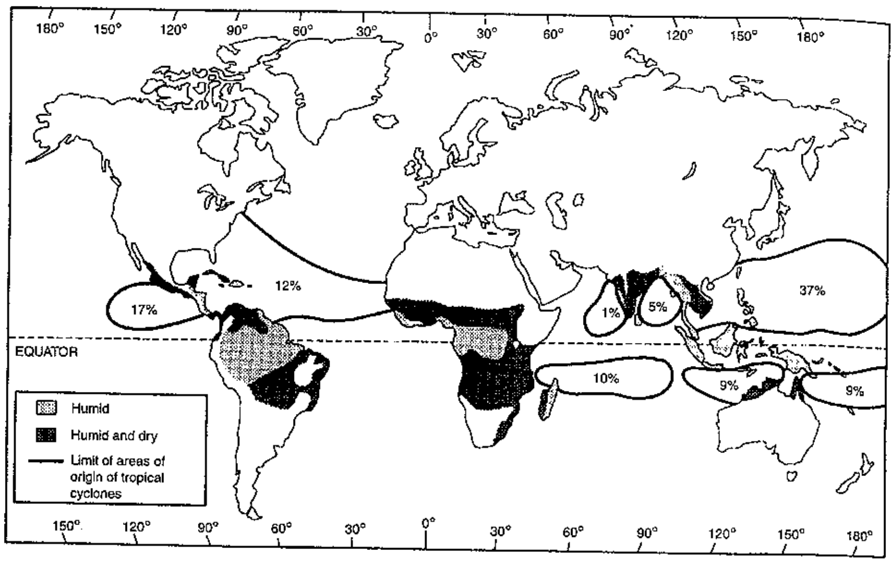


Figure 170 Tropical climates according to the Köppen system (modified by Trewartha), and occurrence of tropical cyclones shown as percentages (After WMO (1983))

development in the different climatic zones. In contrast, the parallel retreat of hillslopes, first proposed to explain some semi-arid landscapes in the USA, was extended by King (1953, 1957) as a universal model for hillslope development. King (1962) applied his ideas to all the Gondwana continents, and by implication, to most of the tropical world. This construct had enormous impact because it proposed a single hypothesis of slope development that was linked to 'continental drift' and the emerging theory of plate tectonics.

King's views contradicted the concept of a 'climatic geomorphology', but had little influence in Germany, where W. Penck (1924) had previously linked the extension of plains to tectonics, in a way that did not conflict with climatic control over geomorphic processes. In particular, the importance of a deeply weathered mantle in explanations of humid tropical relief was central to Büdel's (1957) hypothesis of *Doppelten Einebnungsflächen*, or 'double surfaces of levelling'. This paper led geomorphologists in the tropics to rediscover earlier accounts of weathering and relief development, especially in Africa (Falconer 1911; Wayland 1933; Willis 1936). These authors shared the view that plateau landscapes in the humid tropics were lowered incrementally by stripping and renewal of the SAPROLITE cover. Wayland (1933) called the resulting landscapes 'etched plains' (subsequently *etchplains*), the term later adopted by Thomas (1966, 1994) and Büdel (1982) (Figure 171). Falconer (1911) argued that much of the rocky relief in north Nigeria was due to differential weathering, and Willis (1936) argued that the prominence of granite monoliths or BORNHARDTs was due to repeated cycles of weathering and lowering of more susceptible rocks over geologic time. The isolation of bornhardts by selective weathering was also argued by Rougerie (1955) in Ivory Coast and in Brazil by Birot (1958) who stressed that *les dômes crystallines* were not inselbergs, as often described, but are found in zones of dissection. The importance of DEEP WEATHERING, followed by stripping was advocated by Ollier (1959, 1960) for Uganda, and new attempts were made to document the depth and distribution of the weathered mantle by Thomas (1966).

Linton (1955) drew on ideas about tropical weathering to explain 'the problem of tors' in

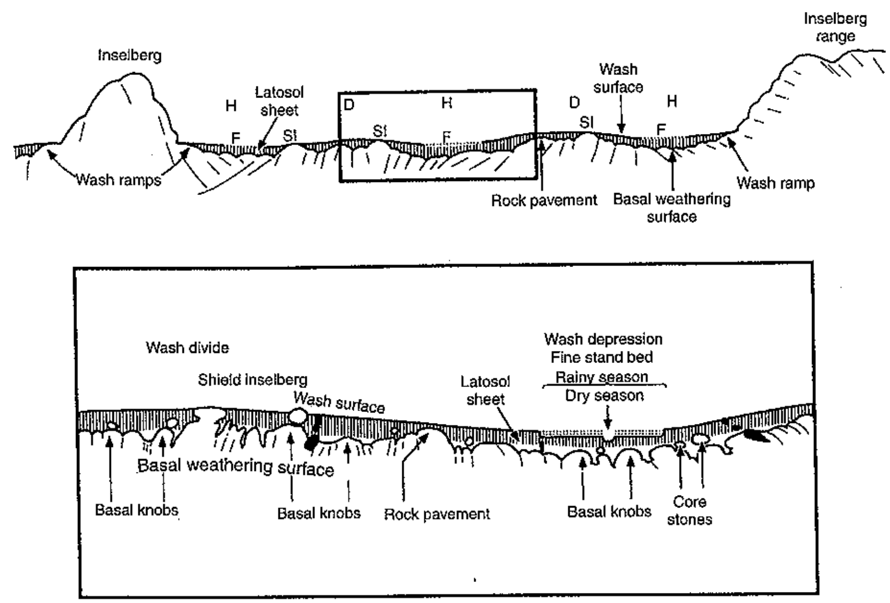


Figure 171 Characteristics of an active etchplain as described by Büdel (1982), based on studies of the Tamilnadu Plain in south India. H: wash depressions; D: wash divides; SI: shield inselbergs; F: fine sand in rainy season riverbed. Details of the wash divide and wash depression are shown in the expanded box. Lowering of the landscape is achieved by rock decay at the basal weathering surface and removal of sands and clays from the wash surface by seasonal runoff

Britain, leading to a clash with King. These ideas were applied to tor landscapes in North America, and to help explain the relief of glaciated areas (Feininger 1971). The notion of changing climates and the existence of 'relief generations' in world landscapes, was reviewed in a European context by Bakker and Levelt (1964), and became essential to a revised 'climato-genetic' geomorphology (Büdel 1968). Stoddart (1969) in an influential review of climatic geomorphology was scathing. He thought that the concept had its roots in Davisian theory, and considered that observations of landform differences between climatic zones were 'methodologically trivial' (p. 213). This view was widely accepted by those who pointed out that the same physical laws apply everywhere, and that the processes of chemical weathering and water flow are similar in all climates.

Although the Davisian paradigm prevailed for half a century, by 1950 proponents of the new 'process geomorphology' were directing attention away from geomorphic cycles in order to understand the mechanisms of landform change, rediscovering the work of G.K. Gilbert (1877, 1914). In the tropics, Branner (1896) had reported from east Brazil the roles of high rainfall and the presence of organic acids in tropical weathering; recognized the involvement of kaolinized saprolite in generating landslides under a natural forest cover, and the importance of soil fauna such as termites. By the 1950s geomorphologists in the tropics had begun to study these processes. White (1949) emphasized the importance of landslides in Oahu (Hawaii), while in West Africa studies of chemical dissolution of rocks, runoff and erosion arose from French research (Corbel 1957; Fournier 1960; Rougerie 1960). In South America, Tricart (1956, 1959) focused attention on fluvial processes, and the impacts of climate change. His 'Le modelé des régions chaudes, forêts et savanes' (Tricart 1965) was the first 'modern' study of tropical geomorphology. Tricart and Cailleux (1965, English trans. 1972) also published their 'Introduction à la géomorphologie climatique' in the same year complaining (pp. xiii) that 'the study of tropical geomorphology...has been delayed by the uniformist ideas of "normal erosion" ', which led to the systematic neglect of the tropics. This neglect was compounded by problems of access, fear of disease and, in the rainforests, a lack of inter-visibility.

By this time development involving new roads and construction projects had exposed the reality and extent of deep weathering in the humid tropics. In Hong Kong, detailed studies of weathering (Ruxton and Berry 1957) and analyses of intense tropical rainfall (Lumb 1975) showed unequivocally the importance of both to an understanding of landslide occurrence. The devastating effects of tropical storms was also reported from Rio de Janeiro (Jones 1973). On the other hand, the high permeability of many tropical soils was documented by Nye (1954, 1955) and subsequently by many others, who found that infiltration rates in *ferrallitic* soils (*oxisols, ultisols*) over gneissic rocks exceeded 200 mm h^{-1}, greater than expected rainfall intensities. Nye did not observe surface runoff in his experiments at Ibadan, Nigeria, but others have measured significant runoff on moderate slopes at intensities of 60–70 mm h^{-1} (Morgan 1986). Outside the equatorial zone, the seasonal concentration of tropical rainfall combined with the high intensity of individual falls convinced many writers of its potential for serious soil erosion (Fournier 1960; Roose 1981). Studies of sediment yield have recorded great variation (Douglas and Spencer 1985; Thomas 1994). On undisturbed plots under rainforest, yields of 0.1–10 t $km^2 yr^{-1}$ have been recorded, but small erosion plots left bare of plant cover can return sediment yields of 5,000 t $km^2 yr^{-1}$. Small catchments in deforested highland areas have recorded from 1,500–3,000 t $km^2 yr^{-1}$, but larger catchments, which have floodplains for major sediment storage, produce far less sediment per unit area and the average for Africa is just 35 t $km^2 yr^{-1}$, rising to 380 t $km^2 yr^{-1}$ in Asia. Rates are more influenced by catchment relief and rainfall amount than by other factors (Milliman and Syvitski 1992).

Tropical rivers are as varied as in other regions. Large rivers traversing the plateaux and plains of the southern (Gondwanaland) continents appear distinctive because of the fine calibre of their bedload (usually sand) and the interruption of their thalwegs by rock outcrops, which create rapids and waterfalls. Opinions formed on the basis of these characteristics hold that the thoroughness of tropical weathering reduces most rocks to sand and clay, or causes their total dissolution. According to writers such as Büdel (1957), Birot (1958) and Tricart (1965) this leads to a lack of abrasive tools for channel cutting; valleys are consequently open with shallow slopes, and river channels are interrupted by resistant bands of rock, which they do not erode effectively. However, wherever tropical rivers traverse hill ranges or fault scarps coarse bedload material becomes abundant and deep valleys and gorges result. Deep valleys are also cut into saprolite.

Great seasonal variations in river discharges are found in monsoon areas, and appear to lead to distinctive channel morphologies. The Auranga River, India was described by Gupta and Dutt (1989) as having a braided sandsheet with shallow (30 cm) channels in the long dry season, but is converted during the monsoon rains (1,500 mm in 4 months) to a wide meandering channel with sand and gravel point bars. By contrast rivers draining the equatorial Papua New Guinea mountains have a very low seasonal variability in flow and are subject to almost instantaneous runoff from steep slopes kept in a near saturated state by frequent rainfalls. Rainfall intensity is lower in this environment and most sediment reaches the river via frequent landslides (Pickup 1984; Pickup and Warner 1984).

Justification for a 'geomorphology of the tropics' lies in the balance of processes and their outcomes in terms of weathered materials and soil properties, erosional forms and sediments. Paradoxically, the major rock landforms, characterized by the granite domes or *bornhardts* that previously received such emphasis, owe more to petrology and structure than to specific climatic parameters. Nonetheless many illustrate the principles of selective deep weathering and stripping of regolith over very long time periods. Although it has been argued that lateritic (ferrallitic) weathering is more a product of time than of climate (Taylor *et al.* 1992), these factors are not independent. Advanced weathering occurs beneath ancient landsurface in non-tropical Australia, but is found more widely in the humid tropics, where some Neogene formations are also affected (Thomas *et al.* 1999). Weathering rates are, however, greatly influenced by water movement

through the profile, so that dry climates lacking a seasonal water surplus have a much reduced potential for deep and advanced weathering. Bourgeon (see Pedro 1997) showed that deep clay-rich kaolinitic 'alterites' pass into shallow, sandy materials (arènes, grus) along an east–west transect of southern peninsular India (P > 2,000 mm y^{-1}–700 y^{-1}), where smectite increases in proportion to kaolinite as the climate becomes drier.

Properties of the saprolite are fundamental to the understanding of tropical soils, as much for engineering as for agriculture. The permeability of tropical saprolite is increased by Fe_2O_3 adhesion to kaolin platelets, which produces larger aggregates. But these form 'cardhouse' fabrics that can collapse under loading, causing settlement of building structures. On the other hand, 2:1 lattice clays (smectites) found in seasonal climates, experience remarkable expansion as water content increases. This reduces permeability and causes 'heave' in many drier areas, forming GILGAI. Iron enrichment of the saprolite can arise from the flux of Fe^{2+} ions that are subsequently oxidized to Fe_2O_3. The Fe is mobilized as Fe^{2+} under conditions of low pH or by chelation involving organo-metallic complexes. This is likely to occur within poorly drained depressions in humid climates or under forest. But the fixation of Fe needs a rise in pH accompanied by drying of the deposit. The term *laterite* is often used to describe materials with high Al/Fe sesquioxide content (see Aleva 1994). It is a form of FERRICRETE that may also form in transported sediments; it becomes indurated and resistant to erosion if subject to wetting and drying, and may appear as hill cappings and valleyside benches in the landscape. Seasonality of rainfall contributes to this process, but in many cases the duricrusts have formed due to falling water tables, resulting from climate change and/or dissection. Beneath well-drained sites in the rainforest Fe does not become concentrated, but Al_2O_3 rich *bauxite* (mainly gibbsite) forms often as irregular nodules. Beneath undisturbed forests, the dominant exports are ions in solution, leading to chemical sediments offshore (Erhart 1955). In equatorial environments, where water tables are high, almost complete dissolution of silicate minerals is possible, leading to the formation of quartz-dominated 'white sands' (Thomas 1994).

In the mid-Miocene, there was a marked reduction of rainfall in the southern tropics, probably associated with the growth of the Antarctic ice sheet. Large sandsheets, such as the Kalahari Sands that penetrate the Congo Basin rainforests from the south may have originated at this time, and it has been argued that weathering systems were 'switched off' by this event, but many equatorial areas were humid throughout the Neogene. Quaternary climate change in the tropics, involved temperature reductions of 5–6 °C at the Last Glacial Maximum (LGM), and lowland areas experienced rainfall reductions of 25–50 per cent, and increased seasonality. Rainforests were converted to deciduous woodlands and/or savanna mosaics over large areas. Many small streams in the tropics left few traces of sedimentation during the LGM; other formerly meandering rivers became braided. Alluvial fans formed along many escarpments, probably due to diminished stream power and increased sediment load. Today, tropical landscapes are sensitive to environmental changes because the thick saprolites and stores of Quaternary sediment are subject to rapid erosion under intense rainfalls, if vegetation is cleared.

Accidents of history led to tropical areas lying beyond the immediate experience of geographers and geologists resident in the temperate zones of Europe and North America. This was not the case for the arid, glacial and periglacial environments, and it has led to persistent neglect of tropical geomorphology. Those who argue against the recognition of rainforests and savannas as distinctive environments for geomorphology fail to recognize the linkages between abundant rainfall, warm soil conditions, the productivity of humid tropical ecosystems, and the formative environments for soils and sediments. Ecologists and pedologists have long recognized that system outcomes depend on rates of biogeochemical cycles: of growth and decay; leaching and accumulation. Under tropical forests, rates of weathering and bioturbation can compensate for soil loss and erosion. Little known biota enter into weathering processes and organic acids can alter classic geochemistry to enhance the mobility of many cations. The frequency of intense rainfall in many areas is greater than in non-tropical climates, and the incidence of cyclones with high wind velocities is regionally important. These characteristics increase the hazards of flooding, inundation and landsliding, even in tectonically quiescent areas.

If geomorphology had developed from a tropical perspective it might have emerged with a different emphasis. But our views are also influenced by the available database, and the most serious problem in relating a geomorphology of the tropics to studies elsewhere remains the paucity of data about conditions in tropical environments.

References

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

Bakker, J.P. and Levelt, Th.W.M. (1964) An inquiry into the probability of a polyclimatic development of peneplains and pediments (etchplains) in Europe during the Senonian and Tertiary Period, *Publication Service Carte Géologique, Luxembourg* 14, 27–75.

Birot, P. (1958) Les dômes crystallines, *Memoires et Documents, CNRS* 6, 8–34.

Branner, J.C. (1896) Decomposition of rocks in Brazil, *Geological Society of America Bulletin* 7, 255–314.

Büdel, J. (1948) Das System der klimatischen Morphologie, *Deutscher Geographen-Tag, Munchen*, 65–100.

——(1957) Die 'doppelten Einebnungsflächen' in den feuchten Tropen, *Zeitschrift für Geomorphologie*, N.F 1, 201–288.

——(1968) Geomorphology principles, in R.W. Fairbridge (ed.) *Encyclopaedia of Geomorphology*, 416–422, New York: Reinhold.

——(1982) *Climatic Geomorphology*, trans. by L. Fischer and D. Busche, Princeton: Princeton University Press.

Corbel, J. (1957) L'érosion climatiques des granites et silicates sous climats chauds, *Revue Géomorphologie Dynamique* 8, 4–8.

Cotton, C.A. (1942) *Climatic Accidents in Landscape Making*, Wellington: Whitcomb and Tombs.

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

Douglas, I. and Spencer, T. (eds) (1985) *Environmental Change and Tropical Geomorphology*, London: Allen and Unwin.

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

Falconer, J.D. (1911) *The Geology and Geography of Northern Nigeria*, London: Macmillan.

Feininger, T. (1971) Chemical weathering and glacial erosion of crystalline rocks and the origin of till, *US Geological Survey Professional Paper* 750-C, C65–C81.

Fournier, F. (1960) *Climat et Érosion: La Relation entre l'Érosion du Sol par l'Eau et les Précipitations Atmosphériques*, Paris: Presses Universitaires.

Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, United States Department of the Interior, Washington, DC.

——(1914) The transport of debris by running water, *United States Geological Survey Professional Paper* 86.

Gupta, A. and Dutt, A. (1989) The Auranga: description of a tropical monsoon river, *Zeitschrift für Geomorphologie, N.F.* 33, 73–92.

Jones, F.O. (1973) Landslides of Rio de Janeiro and the Serra das Araras Escarpment, Brazil, *US Geological Survey Professional Paper* 697.

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

King, L.C. (1957) The uniformitarian nature of hillslopes, *Transactions of the Edinburgh Geological Society* 17, 81–102.

——(1962) *The Morphology of the Earth*, Edinburgh: Oliver and Boyd.

Linton, D.L. (1955) The problem of tors, *Geographical Journal* 121, 470–487.

Lumb, P. (1975) Slope failures in Hong Kong, *Quarterly Journal of Engineering Geology* 8, 31–65.

Milliman, J.D. and Syvitski, P.M. (1992) Geomorphic/tectonic control of sediment discharge to the ocean: the importance of small mountain rivers, *Journal of Geology* 100, 525–544.

Morgan, R.P.C. (1986) *Soil Erosion and Conservation* Harlow: Longman.

Nye, P.H. (1954) Some soil-forming processes in the humid tropics. Part I: A field study of a catena in the West African forest, *Journal of Soil Science* 5, 7–27.

——(1955) Some soil-forming processes in the humid tropics. Part II: The development of the upper slope member of the catena, *Journal of Soil Science* 6, 51–62.

Ollier, C.D. (1959) A two cycle theory of tropical pedology, *Journal of Soil Science* 10, 137–148.

——(1960) The inselbergs of Uganda, *Zeitschrift für Geomorphologie N.F.* 4, 43–52.

Pedro, G. (1997) Clay minerals in weathered rock materials in soils, in H. Paquet and N. Clauer (eds) *Soils and Sediments – Mineralogy and Geochemistry*, 1–20, Berlin: Springer.

Penck, W. (1924) Die Morphologische Analyse, *Geographische Abhandlungen* 2.

Pickup, G. (1984) Geomorphology of tropical rivers I. Landforms, hydrology and sedimentation in the Fly and Lower Purari, Papua New Guinea, *Catena Supplement* 5, 1–18.

Pickup, G. and Warner, R.F. (1984) Geomorphology of tropical rivers II. Channel adjustment, sediment load and discharge in the Fly and Lower Purari, Papua New Guinea, *Catena Supplement* 5, 19–41.

Roose, E.J. (1981) Approach to the definition of rain erosivity and soil erodibility in West Africa, in M. De Broodt and D. Gabriels (eds) *Assessment of Erosion*, 143–151, Chichester: Wiley.

Rougerie, G. (1955) Un mode de dégagement probable de certains dômes granitiques, *Comptes Rendus Academie des Sciences* 246, 327–329.

——(1960) *Le Façonnement Actuel des Modelés en Côte D'Ivoire Forestière*, Memoire Institut Français Afrique Noire 58, Dakar.

Ruxton, B.P. and Berry, L. (1957) Weathering of granite and associated erosional features in Hong Kong, *Geological Society of America Bulletin* 68, 1,263–1,292.

Scrivenor, J.B. (1931) *The Geology of Malaya*, London: Macmillan.

Stoddart, D.R. (1969) Climatic geomorphology: review and assessment, *Progress in Physical Geography* 1, 160–222.

Taylor, G.R., Eggleton, R.A., Holzhauer, C.C., Maconachie, L.A., Gordon, M., Brown, M.C. and McQueen, K.G. (1992) Cool climate lateritic and bauxitic weathering, *Journal of Geology* 100, 669–677.

Thomas, M.F. (1966) Some geomorphological implications of deep weathering patterns in crystalline rocks in Nigeria, *Transactions of the Institute of British Geographers* 40, 73–193.

——(1994) *Geomorphology in the Tropics*, Chichester: Wiley.

Thomas, M.F., Thorp, M.B. and McAlister, J. (1999) Equatorial weathering, landform development and the formation of white sands in Northwestern Kalimantan, Indonesia, *Catena* 36, 205–232.

Thorbecke, F. (ed.) (1927) *Morphologie der Klimazonen*, Breslau: Düsseldorfer Geographische Vorträge.

Tricart, J. (1956) Comparaison entre les conditions de façonnement des lits fluviaux en zone tempérée et zone intertropicale, *Comptes Rendus Academie Sciences* 245, 555–557.

——(1959) Observations sure le façonnement des rapides des rivières intertropicales. *Bulletin Section Géographique, Comité Travaux Historique et Scientifique* 68, 333–343.

——(1965) *The Landforms of the Humid Tropics, Forests and Savannas*, trans. C.J. Kiewiet de Jonge (1972), London: Longmans.

Tricart, J. and Cailleux, A. (1965) *Introduction à la Géomorphologie Climatique*, Paris: SEDES.

Wayland, E.J. (1933) Peneplains and some other erosional platforms, *Annual Report and Bulletin, Protectorate of Uganda Geological Survey, Department of Mines*, Note 1, 77–79.

White, L.S. (1949) Process of erosion on steep slopes of Oahu (Hawaii), *American Journal of Science* 247, 168–186.

Willis, B. (1936) East African Plateaus and Rift Valleys, Studies in Comparative Seismology, *Carnegie Institute, Washington, Publication*, 470.

WMO (World Meteorological Organization) (1983) Operational hydrology in the humid tropical regions, in E. Keller (ed.) *Hydrology of Humid Tropical Regions with Particular Reference to the Hydrological Effects of Agriculture and Forestry Practice*, IAHS Publication, 140, 3–26.

Further reading

Dubreuil, P.L. (1985) Review of field observations of runoff generation in the tropics, *Journal of Hydrology* 80, 237–264.

Faniran, A. and Jeje, L.K. (1983) *Humid Tropical Geomorphology*, Harlow: Longman.

Fookes, P.G. (1997) *Tropical Residual Soils*, The Geological Society, London.

Gutierrez Elorza, M. (2001) *Geomorphología Climática*, Barcelona: Omega.

Taylor, G. and Eggleton, R.A. (2001) *Regolith Geology and Geomorphology*, Chichester: Wiley.

Wirthmann, A. (2000) *Geomorphology of the Tropics*, Berlin: Springer.

MICHAEL F. THOMAS

TROTTOIR

Trottoirs (surf ledges) are narrow, subhorizontal erosional SHORE PLATFORMS with a veneer of Vermetid gastropod tubes (see VERMETID REEF AND BOILER) and encrusting coralline algae, and surfaces that often consist of tiers of VASQUES. Trottoirs occur at, or a little below, mean sea level in tropical seas and in the warmer parts of the southern and eastern Mediterreanean. They have been attributed to CORROSIONal cliff erosion in the spray zone and possibly the protection afforded to the wave-battered seaward edges of the platforms by Vermetid tubes, algae and other organic encrustations. On Curacao and other islands in the southern Caribbean Sea, trottoirs, up to 10 m width, are restricted to exposed areas where there are thick encrustations, and they are replaced by notches (see NOTCH, COASTAL) in less exposed areas. It has been suggested that trottoirs may eventually attain a state of DYNAMIC EQUILIBRIUM with the processes operating on them as their increasing width reduces the rate of cliff erosion.

Further reading

Focke, J.W. (1978) Limestone cliff morphology on Curacao (Netherlands Antilles), with special attention to the origin of notches and vermetid/coralline algal surf benches (corniches, trottoirs), *Zeitschrift für Geomorphologie* 22, 329–349.

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

TSUNAMI

A tsunami is a Japanese word that describes a 'harbour wave' and it is used within the scientific community to describe a series of waves that travel across the ocean with exceptionally long wavelengths (up to several hundred kilometres between the wave crests in the open ocean). As the waves approach a coastline, the speed of the waves decreases as they are deformed within shallower water depths. During this process of wave deformation, the height of the wave increases significantly and as the waves strike the coastline they often cause widespread flooding across low-lying coastal areas and on many occasions cause loss of life and widespread destruction of property. Tsunamis are frequently described in the media as tidal waves. However this view is completely wrong since tsunamis have nothing to do with tides or weather. Most tsunamis in the world occur in the Pacific region. For example, during the past ten years, damaging tsunami floods have taken place in Nicaragua (1991), Flores, Indonesia (1992) and Okushiri Island, Japan (1993). The most recent destructive tsunami disaster took place in Papua New Guinea (1998) where several thousand people lost their lives.

Frequently tsunamis are described as seismic sea waves that are produced as a result of a sudden displacement of part of the seafloor. Usually a seismic disturbance associated with an offshore earthquake will cause rupturing of the ocean floor and when this happens the overlying water column is disturbed. It is during this phase of disturbed motion that tsunamis are often generated. Underwater earthquakes in tectonically active areas of the world (e.g. the Pacific rim) are the most common cause of tsunamis. For example, as recently as May 1960, a large underwater earthquake took place off the coast of Chile. The tsunami travelled across the Pacific Ocean. After twelve hours it struck the coastline of the Hawaiian island chain, while after a further twelve hours it reached the coast of Japan where it destroyed all in its path. A huge tsunami was produced as a result of a large earthquake beneath the seafloor west of Portugal on 1 November AD 1755. One observer who witnessed this tsunami was William Borlase who observed its arrival on the shore of Mounts Bay, Cornwall. He noted:

> the first and second refluxes were not so violent as the third and fourth (*tsunami waves*) at which time the sea was as rapid as that of a mill-stream descending to an undershot wheel, and the rebounds of the sea continued in their full-fury for fully two hours... alternatively rising and falling, each retreat and advance nearly of the space of ten minutes, till five and a half hours after it began.

Another observer of this tsunami described how, at Larmorna Cove, Cornwall:

> the sea on this occasion rushed suddenly towards the shire in vast waves, with such impetuosity, that large rounded blocks of granite from below low-water mark were swept along like pebbles, and many of them deposited far beyond high water mark. One large block, weighing probably 6 or 8 tons, was borne repeatedly to and fro several feet above the level of high water, and at length deposited about ten feet above that level in the stream, where it still lies.

Tsunamis can also be generated by underwater landslides and by landslides occurring above sea level and moving downslope into the sea (see SUBMARINE LANDSLIDE GEOMORPHOLOGY). The occurrence of such slides is quite common on a geological timescale. In 1929, an offshore earthquake led to widespread sediment slumping and the generation of turbidity currents across the seafloor off the Grand Banks of Newfoundland as well as a tsunami that locally reached up to +30 m at the coast. In the North Atlantic region a very large tsunami was generated approximately 8,000 years ago by one of the world's largest underwater landslides (the Second Storegga Slide). The tsunami caused flooding along parts of the Norwegian coastline up to levels +20 m above sea level and along the UK coastline where the highest flood levels reached up to +6 m above sea level. One of the most widely described processes of submarine slide generation is through the release of methane gases (clathrates) contained within ocean-floor sediments. It is believed that the sudden release of such gases can cause local slumping and sliding. The second process arises from low-magnitude earthquakes which in themselves have no destructive effect, but which are of sufficient intensity to induce shaking of seafloor sediments thus causing downslope slumping and sliding of sediment. Generally speaking, underwater landslides generate tsunamis with much less energy than those produced by earthquake-triggered faulting on the ocean floor. The size and energy of landslide-generated tsunamis decreases rapidly with increased distance from the source area.

Occasionally, tsunamis can be generated by the impact of meteorites onto the ocean surface. Perhaps the most famous tsunami associated with a large meteorite impact was that which took place *c*.65 million years ago. The so-called K-T impact, usually associated in the geological literature with the extinction of the dinosaurs, also created a huge tsunami. Traces of the tsunami waves can be found in areas of Mexico and Texas. Most authoritative accounts consider that, given the proportion of the Earth's surface area represented by the Pacific Ocean, a meteorite capable of striking the Pacific Ocean and generating a tsunami is in the order of 1:400,000 years.

Tsunamis can also be produced through the collapse of a volcanic crater into the sea during a major explosive volcanic eruption. Although such a phenomenon may have occurred *c*.18,000 years ago in the case of the volcanic eruption of Santorini in the Aegean Sea, such events are extremely rare. Similarly, tsunamis can be produced as a result of the collapse of the flank of a volcano into the sea. Such tsunamis may be generated by rockfalls, rockslides, rotational slips or debris flows that originate on steep slopes and move into the sea. Such an event happened during the recent eruptions on the island of Montserrat. Some authors have proposed that certain hillslopes on the flanks of some of the now dormant volcanoes in the Canary Isles may have collapsed into the sea in the past and produced large tsunamis.

Reference

Bryant, E. (2001) *Tsunami: The Underrated Hazard*, Cambridge: Cambridge University Press.

ALASTAIR G. DAWSON

TUFA AND TRAVERTINE

Tufa and travertine are terrestrial freshwater accumulations of calcium carbonate, whose formation often involves a degree of organic involvement. The names tufa and travertine can be used synonymously, but often tufa is taken to refer to a softer, more friable deposit whilst travertine refers to a harder, more resistant material frequently used as a building material. Tufa derives from tophus or tufo, used in Roman times to describe crumbly white deposits (Ford and Pedley 1996; Pentecost 1993). Travertine derives from the Latin *lapis tiburtinus*, or Tibur stone, and was originally used to describe the massive, hot spring deposits around Rome (Pentecost 1993). Some travertines are of hydrothermal origin and thus contain very limited plant material, whilst tufa and travertines precipitated by cool water typically contain the remains of micro- and macrophytes, invertebrates and bacteria. Different usages of the terms tufa and travertine are found in different countries, and there is as yet no standard international terminology although Ford and Pedley (1996) and Pentecost and Viles (1994) suggest two alternative schemes. Tufa and travertine are distinguished from CALCRETE, SPELEOTHEMS and STROMATOLITES by their environment of formation, but they share many similar characteristics and may grade into one another in some circumstances.

Tufas and travertines form in freshwater environments where thermodynamic and kinetic characteristics favour the precipitation of calcium carbonate from carbonate-rich waters. Such conditions arise where carbon dioxide is removed from the water through turbulent degassing, evaporation or biological uptake. Suitable conditions for precipitation of tufa and travertine are often found in or near KARST areas where dissolution of limestone provides high levels of dissolved carbonate, or where thermal waters rich in carbon dioxide originate in areas of recent volcanic activity. Despite difficulties in obtaining reliable dates of tufas using ^{14}C, most tufas and travertines which have formed from cool water appear to originate from the late Quaternary to early Holocene. This is certainly true in Britain and much of Europe, although in some parts of the world such tufas and travertines are still forming rapidly today. There has been much debate over the major controls on tufa and travertine formation and in particular the role of organisms in their genesis. Certainly, aquatic plants and micro-organisms can aid deposition of tufa – by providing precipitation nuclei, by removing carbon dioxide from the water and perhaps also by direct precipitation of calcium carbonate – but in some environments physico-chemical controls on precipitation outweigh any biological involvement. In many fluvial tufas within forested catchments inorganic modes of precipitation dominate in the turbulent upper reaches, whilst further down tree debris contributes to barrage architecture and macro- and microphytes, and even insect larvae can aid deposition (Drysdale 1999). Hydrothermal travertines are precipitated through largely inorganic processes, although Robert Folk (1994) has suggested that tiny nannobacteria may also play a key role.

Tufa and travertine are found on all continents except Antarctica and there are several very impressive deposits around the world, such as the barrages of the Plitvice Lakes in Croatia, the travertine complex of Antalya in south-west Turkey (Burger 1990) and the Huanglong ravine terraces in north-west Sichuan, China. Good summaries of the occurrence and nature of the major tufa and travertine deposits in Britain, Europe, China and the world are provided by Pentecost (1993), Pentecost (1995), Pentecost and Zhaohui (2001) and Ford and Pedley (1996) respectively. Many occurrences of large tufa and travertines have still not yet been properly documented in the international literature, such as the extensive deposits within the arid Naukluft area of south-central Namibia (Plate 141).

Tufas and travertines form in a range of different geomorphic settings, including fluvial, lacustrine, paludal and spring environments. Within rivers, tufas and travertines can form spectacular barrages often with waterfalls cascading over them, and with clastic tufa accumulating behind the barrage. In some fluvial environments, suites of lakes become created between barrages. Much smaller accumulations of tufa and travertine also occur in many streams, producing fluvial crusts and oncoids. In lacustrine environments, similar oncoids and crusts occur as well as larger reef-like accumulations. In marshy environments, low

'late 141 A large cone-shaped tufa, *c.*400 m across and nearly 100 m high, developed on the edge of the Naukluft Mountains at Blasskrantz in central Namibia. It has formed where a seasonal stream cascades over a steep slope, permitting degassing to take place

late 142 A small tufa dam formed across a creek in the tropical Napier Range of the Kimberley district, northwestern Australia

relief muddy tufas tend to develop, often mixed with marls and chalks. Around springs, mounds and terraces can develop, and where springs debouch on steep slopes such deposits can form huge prograding cascades. Tufas and travertines can produce great geomorphological change within a catchment, as they influence river flows through the creation of barrages and can armour slopes and change the course of springs.

The fabric of tufas and travertines reveals much about their mode of formation, and can also contain useful palaeoenvironmental information. Where organic influence in tufa and travertine precipitation is debated, the petrology of the deposit can provide helpful insights. Martyn Pedley has provided a useful concept of tufa and travertine as phytoherms, with different facies found in different parts of the deposits and a clear organic role played by biofilms in the formation of many zones (Pedley 1992). Some tufas and travertines possess clear laminations which can be of great use in environmental reconstruction. Tufas precipitated in association with cyanobacteria from the genus Phormidium for example, tend to display seasonal banding with alternate sparitic, light layers (representing inorganic deposition in autumn and winter) and micritic, dark layers (formed as a result of cyanobacterially mediated deposition in spring and summer).

The fact that many major cool water tufas and travertines date from the late Quaternary and early Holocene has led several authors to conclude that there is a strong climatic influence on their formation (see for example Goudie *et al.* 1993). Alternatively, or as well as such climatic control, human impacts may also be responsible for a recent decline in the accumulation of tufas and travertines in some parts of the world. Isotopic and trace element contents of tufas and travertines can be used to reconstruct palaeoenvironmental conditions within different parts of dated sequences, thus throwing light on both the environmental conditions and their influence on tufa deposition rates. Andrews *et al.* (1997) illustrate the utility of stable isotopes of oxygen and carbon in reconstructing palaeoenvironmental conditions in Holocene tufa deposits, and Matsuoka *et al.* (2001) show how trace element contents within laminated tufas can provide high resolution records of climate and catchment conditions. As Ford and Pedley (1996) conclude 'tufas have an untapped potential to provide the best land-based opportunity for accessing shorter-term Holocene environmental change'.

References

Andrews, J.E., Riding, R. and Dennis, P.F. (1997) The stable isotope record of environmental and climatic signals from modern terrestrial microbial carbonates from Europe, *Palaeogeography, Palaeoclimatology, Palaeoecology* 129, 171–189.

Burger, D. (1990) The travertine complex of Antalya/Southwest Turkey, *Zeitschrift für Geomorphologie Supplementband* 77, 25–46.

Drysdale, R.N. (1999) The sedimentological significance of hydropsychid caddis-fly larvae (order: Trichoptera) in a travertine-depositing stream: Louie Creek, Northwest Queensland, Australia, *Journal of Sedimentary Research* 69, 145–150.

Folk, R.L. (1994) Interaction between bacteria, nannobacteria and mineral precipitation in hot springs of central Italy, *Geographie physique et Quaternaire* 48, 233–246.

Ford, T.D. and Pedley, H.M. (1996) A review of tufa and travertine deposits of the world, *Earth-Science Reviews* 41, 117–175.

Goudie, A.S., Viles, H.A. and Pentecost, A. (1993) The Late Holocene tufa decline in Europe, *Holocene* 3, 181–186.

Matsuoka, J., Kano, A., Oba, T., Watanabe, T., Sakai, S. and Seto, K. (2001) Seasonal variation of stable isotopic compositions recorded in a laminated tufa, SW Japan, *Earth and Planetary Science Letters* 192(1), 31–44.

Pedley, H.M. (1992) Freshwater (phytoherm) reefs: the role of biofilms and their bearing on marine reef cementation, *Sedimentary Geology* 79, 255–274.

Pentecost, A. (1993) British travertines: a review, *Proceedings, Geologists' Association* 104, 23–39.

——(1995) Quaternary travertine deposits of Europe and Asia Minor, *Quaternary Science Reviews* 14, 1,005–1,028.

Pentecost, A. and Viles, H.A. (1994) A review and reassessment of travertine classification, *Géographie physique et Quaternaire* 48, 305–314.

Pentecost, A. and Zhaozhui, Z. (2001) A review of Chinese travertines, *Cave and Karst Science* 28, 15–28.

HEATHER A. VILES

TUNNEL EROSION

Tunnel erosion is an insidious form of land degradation that is initiated in subsoil and/or substrata and remains inconspicuous until considerable damage has occurred. This type of water erosion is found in earthworks as well as hillslopes and in the latter case refers to the hydraulic removal of subsurface material causing the formation of underground channels in the natural landscape (Boucher 1990). Tunnelling takes place primarily when the shear stress applied by flowing water enlarges an existing macropore or passageway (Bryan and Jones 1997). Corrasion of the tunnel perimeter by ensuing flow and the impact of vehicular, livestock and/or human traffic cause localized collapse of the land surface, producing sinkholes which can entrain surface runoff. The natural bridges of soil remaining between sinkholes are destroyed when the cavities merge, leaving a gully. Debris fans are most conspicuous on hillslopes, beginning from points on the land surface where the translocated sediment is debouched, and consisting of relatively coarse material which can no longer be transported by the hydraulic head. Finer particles are washed further downslope. The initial length of tunnels is uncertain but they may be a series of interconnected macropores and can extend to several hundred metres. The diameter of these features typically ranges between several centimetres and a few metres, the latter proving to be most frequent in semi-arid environments.

Tunnelling has been recorded under a wide range of climates in many different materials on landscapes subjected to diverse land-use histories, but owing to occurrence in uninhabited forest in Papua New Guinea and Eocene palaeopiping in the USA, it should be seen as a naturally occurring process. Sources of tunnelflow include rainfall, snowmelt, irrigation water and ground water. Macropores, desiccation and structural cracks, surface depressions, decayed tree roots and the burrows of insects, crabs, moles, rodents and rabbits have been shown to be points where surface water can infiltrate directly to the subsoil. The materials eroded have included soils that remained highly stable on wetting in the UK (Jones 1981), glaciolacustrine silts and permafrost in Canada as well as ignimbrite and pumice in New Zealand. Many reports of tunnelling in semi-arid climates, especially southeastern Australia, have been associated with nonsaline sodic texture-contrast soils that slake and/or disperse readily on contact with water. An important characteristic of tunnel erosion is the requirement for a layer(s) of material of relatively low permeability which act(s) as a barrier to further vertical percolation. This material can be breached by cracks, joints, tree roots and well-connected macropores. A hydraulic gradient is required to generate flow and the most common outlets occur on the hillslope proper where a hydraulic head can be seen, or comprise various types of free faces such as gully walls and stream banks.

Whilst most hydrologic research has been conducted in humid climates where ephemeral, seasonal and perennial systems are observed (e.g. Wales, UK, Jones 1981, Bryan and Jones 1997), few data are available for semi-arid environments. On grazing land in Victoria, Australia, flow in a shallow tunnel system (generally less than 1 m deep from the soil surface) responded rapidly to rainfall, and it was clear from the typically short lags between peak rainfall and peak runoff that the soil matrix had been largely bypassed. The recession was also rapid and ground water was not a component of the hydrograph (Boucher 1995). Similar characteristics were documented for shallow tunnels in an area of badlands in western Canada (Bryan and Harvey 1985) and a loess plateau in northern China. However, at least partly owing to internal collapse, the relations were more complex for deep-seated systems which were 20 m to 30 m deep in the badlands area, whilst the inlets ranged between less than 0.5 m and over 20 m in both depth and diameter in the latter catchment. The proportion of discharge passing through tunnels to the stream was estimated as up to 10 per cent and 80 per cent respectively. Therefore, tunnel erosion can be a rapid and significant form of soil and water loss, and the economic implications need investigation. Generally, reclamation of land in southeastern Australia should combine deep ripping of soil to destroy the established flowpaths, chemical amelioration of dispersive soils with gypsum in order to displace the sodium and generate an electrolyte effect, and revegetation with grasses which use up excess water.

Bryan and Jones (1997) suggested that tunnelling and piping are distinct processes which are often difficult to distinguish in practice, and that all subsurface erosion is usually combined as piping. However, from research in Australia and China it appears that the terms are still used interchangeably.

References

Boucher, S.C. (1990) *Field Tunnel Erosion: Its Characteristics and Amelioration*, Clayton, Australia: Monash University and East Melbourne: Department of Conservation and Environment.

——(1995) Management options for acidic sodic soil affected by tunnel erosion, in R. Naidu, M.E. Sumner and P. Rengasamy (eds) *Australian Sodic Soils – Distribution, Properties and Management*, 239–246, East Melbourne: CSIRO Australia.

Bryan, R.B. and Harvey, L.E. (1985) Observations on the geomorphic significance of tunnel erosion in a semi-arid ephemeral drainage system, *Geografiska Annaler* 67A, 257–272.

Bryan, R.B. and Jones, J.A.A. (1997) The significance of soil piping processes: inventory and prospect, *Geomorphology* 20, 209–218.

Jones, J.A.A. (1981) *The Nature of Soil Piping – a Review of Research*, British Geomorphological Research Group Research Monograph No. 3, Norwich: Geo Books.

Further reading

Blong, R.J. (1965) Subsurface water as a geomorphic agent with special reference to the Mangakowhiriwhiri catchment, *Auckland Student Geographer (NZ)* 1, 82–95.

Carey, S.K. and Woo, M-K. (2000) The role of soil pipes as a slope runoff mechanism, subarctic Yukon, Canada, *Journal of Hydrology* 233, 206–222.

Pickard, J. (1999) Tunnel erosion initiated by feral rabbits in gypsum, semi-arid New South Wales, Australia, *Zeitschrift für Geomorphologie N.F.* 43, 155–166.

Slaymaker, O. (1982) The occurrence of piping and gullying in the Penticton glacio-lacustrine silts, Okanagan Valley, BC, in R.B. Bryan and A. Yair (eds) *Badland Geomorphology and Piping*, 305–316, Norwich: Geo Books.

Zhu, T.X., Luk, S.H. and Cai, Q.G. (2002) Tunnel erosion and sediment production in the hilly loess region, north China, *Journal of Hydrology* 257, 78–90.

SEE ALSO: pipe and piping

STUART C. BOUCHER

TUNNEL VALLEY

Tunnel valleys are examples of large-scale, erosional glacial landforms that are formed subglacially. Morphologically, tunnel valleys are overdeepened, elongate depressions with steep, often asymmetric sides, cut into bedrock or unconsolidated glacigenic sediment. A characteristic feature is an undulating long profile, which climbs over bedrock rises and contains overdeepened areas along its floor. Long reaches of the flow path can be against the regional gradient. In dimensions, tunnel valleys can reach up to 4 km in width and over 100 km in length, and depths of erosion can be as great as 400 m below sea level. Commonly they trend oblique to the modern drainage, and many contain subglacial landforms such as drumlins, eskers or gravel dunes. In plan view, tunnel valley systems vary from individual, straight segments, to integrated, anastomosing networks of sinuous valleys. Sedimentary infills of tunnel valleys are diverse, and a wide variety of sediments associated with different depositional environments (glaciterrestrial, glacimarine, glacilacustrine and temperate) have been recorded, which reflect changing conditions during, and subsequent to, valley formation. Sediment gravity flow deposits and glacifluvial sands are particularly common.

There are three main theories of tunnel valley genesis. The first argues that they form time-transgressively, by subglacial meltwater erosion during deglaciation, at or close to the ice margin. The second theory also interprets tunnel valleys as the product of subglacial meltwater erosion but argues that the discharges involved were catastrophic channelized floods, and that tunnel valleys within anastomosing networks form synchronously. Finally, the third theory argues that tunnel valleys cut into unconsolidated sediment are due to creep of deformable sediment into a subglacial conduit from the sides and from below. This material is then removed through the conduit by subglacial meltwater.

Further reading

Ó Cofaigh, C. (1996) Tunnel valley genesis, *Progress in Physical Geography* 20(1) 1–19.

COLM Ó COFAIGH

TURBIDITY CURRENT

Turbidity currents or flows are a type of density flow in which movement on a slope occurs due to changed density between a local fluid and a surrounding fluid (Simpson 1987). In a turbidity current it is the suspended particles that cause the density of the flow to be greater than that of the surrounding fluid. The result of this is that the turbulent suspension moves down any local or regional gravity slope.

Turbidity currents in water can originate in a number of ways. One mechanism is for them to originate as sediment slides and slumps caused by scarp or slope collapse (e.g. when a slope is disturbed by earthquake shocks). Another mechanism is direct underflow of suspension-charged river water in so-called hyperpychnal plumes. These can occur during snowmelt floods in steep-sided FJORDS, in front of RIVER DELTAS, and in river tributaries whose feeder channels have extremely high loads of suspended sediments. A third mechanism is the collection of sediment by longshore drift in the nearshore heads of SUBMARINE VALLEYS and canyons (Leeder 1999).

Turbidity flows are not restricted to water. As powder snow avalanches demonstrate, for example, dense suspensions in air may flow downslope.

Submarine turbidity currents can be up to several kilometres wide and several hundreds of metres thick. They can travel as far as 4,000–5,000 km. Erosion can occur at the base of a flow, producing various scour or sole marks including mega-flutes. Deposition from a waning turbidity current produces sediment accumulations called turbidites. Associated with these may be large-scale asymmetric gravel waves and macrodunes and channel levees (Stow 1994).

References

Leeder, M. (1999) *Sedimentology and Sedimentary Basins. From Turbulence to Tectonics*, Oxford: Blackwell Sciences.

Simpson, J.E. (1987) *Gravity Currents: In the Environment and the Laboratory*, Chichester: Ellis Horwood / Wiley.

Stow, D.A.V. (1994) Deep sea processes of sediment transport and deposition, in K. Pye (ed.) *Sediment Transport and Depositional Processes*, 257–291, Oxford: Blackwell Scientific.

A.S. GOUDIE

TURF EXFOLIATION

A denudation process that is particularly prev... in periglacial areas and leads to the destruc... a continuous vegetation cover thr... removal of soil exposed along small ... Among the processes that lead, p... d tically, to this phenomenon ... ast krake) action, desiccation ... gher zoogeomorphic activitie... s the appears to be especial... imatic regions, not least i... ney are

Reference

Grab, S... Dra... ...

...ustralia's ...oastland. ...l types of ...l *manifestly* ...ers within a ... the meanders ...est underfits can ... air photographs ...at is when a mani... ...b-parallel scarps of ...ormations have been ... In the *underfits of the* ...ing curved around the

A sea... upon w... Turloughs

carboniferous age limestones of central and western Ireland, with the term derived from the Irish *tuar loch* meaning dry lake. In times of decreasing water table (dry seasons) the level of the turlough drops, with the contents draining through the porous parts of the basin floor and via connections to swallow holes. In contrast, times of high rainfall (wetter seasons) will produce a high lake level. Theories of development (see Coxon 1986) are based upon interactions of a karstic landscape formed in the Tertiary and subsequent glaciations, with turlough form dependent upon this history. The flooding regime in turloughs varies considerably, as a result of size, depth, local water table, tidal regime and soil conditions. They can be composed of sand, clay, silt, diamicton, peat or marl, or various combinations (Coxon 1986). An example of a turlough is the Carren Turlough of the Burren Region, County Clare, Eire.

Reference

Coxon, C. (1986) *A Study of the Hydrology and Geomorphology of Turloughs*, Unpublished Ph.D. Thesis, University of Dublin, Ireland.

Further reading

Sweeting, M.M. (1972) *Karst Landforms*, London and Basingstoke: Macmillan.

STEVE WARD

U

UNDERFIT STREAM

A stream is underfit when it is much smaller than the size of its valley. Dury (1964) defines this fluvial geomorphological feature as that of a present-day stream flowing in an alluvial plain and describing free meanders far less ample than those of the enclosing VALLEY MEANDERS, which are frequently ingrown meanders of former large streams. Dury (1965) found that the ratio between the bed width of former and present-day channels (*W/w*) gives an index of underfitness, and that this ratio averages 10/1 in lowland England and near the former ice fronts in Wisconsin (USA), 5/1 and about 3/1 in the Ozarks. He also proposed an alternative index, that given by the wavelength ratios (*L/l*).

Because there is a relationship between stream size and size of meanders, it has to be assumed that some cause has operated to reduce the present stream size significantly below its former values. Davis (1913) had defined underfit streams and related their origin to the capture process taking place during the competitive development of rivers. The beheaded stream underwent shrinkage and reduced the size of the meanders which it had formerly processed. Other hypotheses proposed by subsequent researchers include: underflow through alluvium, deep percolation or cessation of meltwater discharge, overspill from glacial lakes or glacial meltwater and tidal scour. During the 1950s Dury undertook a series of field investigations of subsurface bottoms of valleys occupied by manifest underfits in several localities of the English plain in order to verify the hypothesis that, if valley bends are authentic former meanders, former large channels should have been associated with them. He demonstrated the presence of large meandering channels, up to ten times as wide as the existing channel, cut in bedrock and complete with pool and riffle sequences. Further tests on deep buried channels in the driftless area of Wisconsin, in France and elsewhere confirmed his hypothesis, that there is a sensible constancy of the wavelength ratio between valley meanders and stream meanders throughout entire regions. He rejected previous explanations on the grounds that they could only explain a fraction of the total shrinkage and they were of restricted areal application. He concluded that the required explanation of underfitness had to be a widespread one and, in order to meet this requirement, it could be no other than climatic change. In addition, Dury (1977) found that the RIVER CAPTURE hypothesis could be rejected by means of hydrologic empirical power-functional equations which relate discharge to drainage area.

Underfit streams have now been identified widely in western Europe, where perhaps at least 50 per cent of the length of second and higher order streams is underfit, and as far east as the Ukraine. They occur in all the major climatic regions of the USA, including Alaska. They are present also in the coastal drainage of Australia's Northern Territory and on the eastern coastland.

Dury (1964) distinguished several types of underfit streams (see Figure 172). A *manifestly underfit* is a stream which meanders within a more amply meandering valley, the meanders being of the ingrown type. Manifest underfits can be identified immediately from air photographs or from reliable maps. A *variant* is when a manifest underfit is enclosed by sub-parallel scarps of rock, below which weaker formations have been eroded into a broad trough. In the *underfits of the Osage type*, except for being curved around the

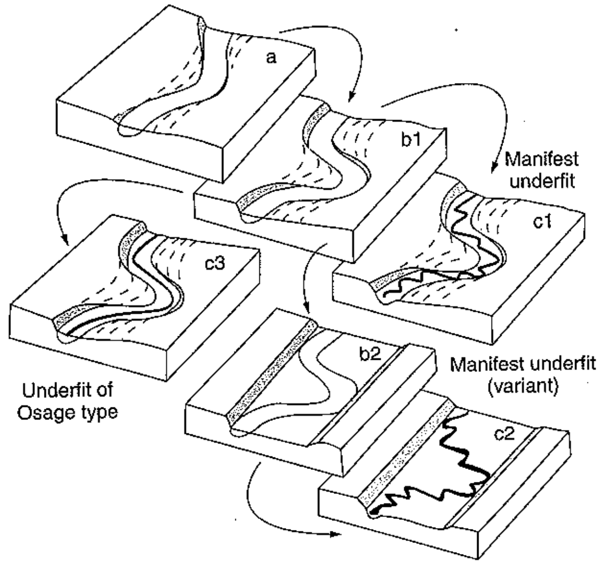


Figure 172 Partial model for the development of underfit streams

valley bends, the channel behaves as if it were straight, but the spacing of the pool and riffle sequences are more closely spaced than the valley meander wavelength would suggest. Osage-type underfits can only be suspected from their plan dimensions and have to be demonstrated from subsurface drilling of the pool and riffle sequence. Although they are named after the Osage River in Missouri, USA, they were first identified by Dury (1966) in the Colo River in New South Wales, Australia.

The development of underfit streams has two aspects: the date of origin of the rivers by which valley meanders have been cut, and the date of the shrinkage by which rivers were reduced to the underfit state. The latest date of origin for ingrown valley meanders is, in many areas, the time when incision of plateaux began. It is known that the incised meandering valleys of the Alpine Foreland and of the Hercynian massifs of Europe were initiated early in the Pleistocene. The high former discharges responsible for shaping valley meanders did not operate throughout the whole interval, but repeated episodes of shrinkage took place. Dates obtained by pollen analysis for the last major streams which are now underfit fall between about 12,000 and 10,000 BP. Evidence from river terraces also places the last main shrinkage well within last-deglacial time, when pluviosity markedly decreased. This appears to have been the time when general underfitness was finally confirmed, except in areas which were still covered by ice or by proglacial lakes.

There is a relationship between stream size and size of meanders. Meander wavelengths of valleys have been estimated to require a BANKFULL DISCHARGE of about 25 times greater than the present, and even 50 or 60 times greater when differences in channel slope, cross section and velocity are incorporated into the estimates. Stream shrinkage is then the result of a significant reduction in discharge at the recurrence interval corresponding to channel-forming flow or to the most probable annual flood. It was suggested by Dury that the channel-forming discharges required to explain the former channel patterns of streams which are now underfit could have been provided by increases in mean annual precipitation of 50 to 100 per cent. Magnitude-area-intensity analysis of precipitation shows that the means of reaching the required increase in precipitation is an increase in the frequency and power of storms.

References

Davis, W.M. (1913) Meandering valleys and underfit rivers, *Annals of the Association of American Geographers* 3, 3–28.
Dury, G.H. (1964) Principles of underfit streams, *US Geological Survey Professional Paper* 452-A.
——(1965) Theoretical implications of underfit streams, *US Geological Survey Professional Paper* 452-C.
——(1966) Incised valley meanders on the Lower Colo River, NSW, *Australian Geographer* 10, 17–25.
——(1977) Underfit streams: retrospect, perspect and prospect, in K.J. Gregory (ed.) *River Channel Changes*, 281–293, Chichester: Wiley.

SEE ALSO: bankfull discharge; river capture; valley meander

MARIA SALA

UNDRAINED LOADING

If a soil is loaded very quickly it can result in there being no time for the drainage of pore water. There may be no consequential change in volume, but pore-water pressures change and result in differential shear and normal stress at every point in the loaded material. This rapidly reduces resistance and sometimes initiates shear movement, or accelerates movement downslope. The eventual dissipation of surplus pore pressure should reinstate steady-state equilibrium in the soil. The extent of pore-water pressure change will vary from soil to soil, as it is predominantly dependent on the soil's composition and its properties. Undrained loading is therefore important in slope stability analysis, particularly for soils with low permeability and receiving rapid loading, and has been incorporated into several modelling studies (e.g. Baker *et al.* 1993). However, such conditions are hard to accommodate within models, as influential factors such as changes in PORE-WATER PRESSURE are difficult to predict.

Reference

Baker, R., Fryman, S. and Talesnick, M. (1993) Slope stability analysis for undrained loading conditions, *International Journal for Numerical and Analytical Methodology in Geomechanics* 17, 15–43.

STEVE WARD

UNEQUAL SLOPES, LAW OF

States that slopes will behave differently depending on their declivities (inclination). The law was proposed by G.K. Gilbert (1877: 140) within his paper on the geology of the Henry Mountains, USA. As rain falls on a slope, the amount of work it can do is proportional to the declivity of the slope. The steeper slope is always degraded faster, and will carry the divide towards the gentler slope. Thus, unless there are equal slope declivities, with homogenous material and identical rainfall, unequal slope activity will proceed. Eventually, a state of equilibrium will form (slope symmetry). Gilbert used this law to explain the form of BADLANDS, alongside his law of divides. This states that slopes on one side of a ridge are independent, while the law of equal declivities establishes a relationship between the slopes, the ridge crest, and the other side of the ridge. Thus, the slopes of the whole ridge behave independently over time and a landscape of unequal slopes develops.

Reference

Gilbert, G.K. (1877) *Geology of the Henry Mountains*, Washington, DC: US Geographical and Geological Survey of the Rocky Mountain Region, 140–141.

STEVE WARD

UNICLINAL SHIFTING

The gradual lateral shifting of a stream or river down-dip as a result of the slope of the underlying bedrock. When a river passes over a zone of inclined alternating hard and soft strata in a valley, it is usually easier for the channel to follow the strike of the less resistant strata, rather than to cut down into the harder rock. This may then induce a lateral shift in the channel. However, the mechanism by which uniclinal shifting occurs remains vague. Differential erosion, and rock permeability have been suggested as important factors, alongside the initial orientation of the channel. An example of uniclinal shifting is found in the Middle Thames Valley, England. Here an extensive flight of aggradation terraces exist on the northern side cascading southwards towards the London Basin syncline, yet there is almost nothing on the southern side that records the river's former route. This suggests the lateral 'uniclinal shifting' of the channel southwards, believed to have occurred in the Pleistocene (Bridgland 1985). Alternative terms to uniclinal shifting are down-dip shifting and 'down-dip migration'.

Reference

Bridgland, D.R. (1985) Uniclinal shifting: a speculative reappraisal based on terrace distribution in the London Basin, *Quaternary Newsletter* 47, 26–33.

STEVE WARD

UNIFORMITARIANISM

Uniformitarianism is a mode of thought which was conventionally defined as 'the present is the key to the past' (Geikie 1905). The names of James Hutton and Charles Lyell are permanently associated with having brought uniformitarian thinking into the main stream of geology, in opposition to the earlier catastrophist thinking, associated with names like Georges Cuvier and Abraham Werner. Prevailing schools of thought in the late eighteenth century had two main tenets: (a) the general belief that God has intervened in history, which therefore has included both natural and supernatural events and (b) the particular proposition that Earth history consists in the main of a sequence of major catastrophes, usually considered as of divine origin. Uniformitarianism as expressed by Hutton (1788) embodied two propositions that were contradictory to these catastrophist views: (a) Earth history can be explained in terms of natural forces still observable as acting today and (b) Earth history has not been a series of universal or quasi-universal catastrophes but has in the main been a long, gradual development. The most obvious example of the confrontation between catastrophists and uniformitarians at the time is provided by the contradictory views on the origin of valleys. Valleys gradually formed by rivers still eroding the valley bottoms were juxtaposed against valleys that had opened up as clefts through divinely controlled revolution. Although this dichotomy made much sense at the time (late eighteenth and early nineteenth century) (Gillispie 1960), there is considerable controversy over the use of the term uniformitarianism at present (Goodman 1967; Shea 1982; Schumm 1991). Hooykaas (1970) says that the use of uniformitarianism should be restricted to a view that states that geological forces of the past differ neither in kind nor in energy from those now in operation. The past should be reconstructed on the assumption that all geological causes of the past were of the same kind and intensity as those of the present.

Gould (1967) subdivided the confusing issues surrounding the definition of uniformitarianism into two components:

(a) a substantive uniformitarianism, which postulates uniformity of kinds and rates of processes (Hooykaas 1970) and
(b) a methodological uniformitarianism comprising a set of two procedural assumptions which are basic to historical enquiry in any empirical science: the principle of the uniformity of natural laws and the principle of simplicity.

Because (a) cannot possibly be true, and was not what Lyell had in mind (Kennedy 2000) when he popularized the term, and because (b) is standard procedure for historical science, the substantive content of the term is superfluous.

Shea (1982) identified twelve fallacies associated with the use of the term uniformitarianism: that uniformitarianism (a) is unique to geology; (b) was first conceived by James Hutton; (c) was named by Charles Lyell, who established its definitive modern meaning; (d) should be called actualism because it refers to the actual or real events and processes of Earth history; (e) holds that only currently acting processes operated during geologic time; (f) holds that the rates or intensities of processes are constant through time; (g) holds that only gradual, non catastrophic processes have occurred during Earth's history; (h) holds that conditions on Earth have changed little through geologic time; (i) holds that Earth is very old; (j) is a theory or hypothesis and can be tested; (k) applies only as far back in history as present conditions existed and only to Earth's surface or crust; (l) holds that the laws governing nature are constant through space and time. He recommends abandoning the term.

Kennedy's definition is, in the last analysis, the most satisfying

> uniformitarianism is a practical tenet held by all modern sciences concerning the way in which we should choose between competing explanations of phenomena. It rests on the principle that the choice should be the simplest explanation which is consistent both with the evidence and with the known or inferred operation of scientific laws. Uniformitarianism is therefore applicable to both historical inference and to prediction of the future outcome of the operation of natural processes (Goodman 1967).
>
> (Kennedy 2000: 502)

References

Geikie, A. (1905) *Founders of Geology*, London: Macmillan.
Gillispie, C.C. (1960) *The Edge of Objectivity: An Essay in the History of Ideas*, Princeton: Princeton University Press.
Goodman, N. (1967) Uniformity and simplicity, *Geological Society of America Special Paper* 89, 93–9.
Gould, S.J. (1967) Is uniformitarianism useful? *Journal of Geological Education* 15, 149–150.
Hooykaas, R. (1970) *Catastrophism in Geology: Its Scientific Character in Relation to Actualism and Uniformitarianism*, Amsterdam: North-Holland Publishing.
Hutton, J. (1788) Theory of the Earth, *Royal Society of Edinburgh Transactions* 1, 209–304.
Kennedy, B.A. (2000) Uniformitarianism, in D.S.G. Thomas and A. Goudie (eds) *The Dictionary of Physical Geography*, 502–504, Oxford: Blackwell.
Schumm, S.A. (1991) *To Interpret the Earth: Ten Ways to be Wrong*, Cambridge: Cambridge University Press.
Shea, J.H. (1982) Twelve fallacies of uniformitarianism, *Geology* 10, 455–460.

Further reading

Simpson, G.G. (1963) Historical science, in C.C. Albritton (ed.) *The Fabric of Geology*, 24–48, Stanford, CA: Freeman, Cooper.

SEE ALSO: actualism; catastrophism; neocatastrophism

OLAV SLAYMAKER

UNIVERSAL SOIL LOSS EQUATION

The Universal Soil Loss Equation is a method for estimating annual SOIL EROSION on the basis of soil loss from a field or hillslope. It was empirically derived from data collected over a twenty-year period from runoff plots at experimental stations established in the 1930s in the United States by the Soil Conservation Service under H.H. Bennett. The object was to measure soil erosion rates under natural rainfall on different soils, slope conditions, cropping and tillage practices, as a basis for SOIL CONSERVATION recommendations. Eventually data were available for twenty-three soils between the Rocky Mountains and the US east coast. Continuing attempts to develop a reliable equation to predict soil erosion culminated in the USLE in 1958 (Wischmeier *et al.* 1958).

The metric version of the equation is:

$$E = R \cdot K \cdot L \cdot S \cdot C \cdot P$$

where E is mean annual soil loss (t ha^{-1}), R is annual rainfall erosivity ($10^7 J ha^{-1}$), K is soil erodibility (relative to a control soil without vegetation cover), L is slope length (relative to a standard slope length of 22.6 m), S is slope gradient (relative to a standard 9 per cent slope), C is crop management (relative to a cultivated bare field), and P is a conservation practices factor (relative to a bare surface without conservation measures).

The most complex and critical factor is annual rainfall EROSIVITY, based on regression analysis of rainfall characteristics to determine those most strongly correlated with soil loss from the runoff plots. The most effective measure is a composite measure involving the total kinetic energy (E; J m^2) during a rainstorm and the maximum rainfall intensity recorded over a 30-minute period during the storm (I_{30}; mm h^{-1}). Annual rainfall erosivity is the sum of EI_{30} for all storms during a year, divided by 1,000. Calculations should be based on records spanning at least twenty-five years, but there are not many locations where such long-term records of rainfall intensity exist. Available data show the highest values from humid tropical areas like the Gold Coast of West Africa, where erosivity exceeds 1,700 (Roose 1977), and the lowest values in temperate and arid regions.

Other data were obtained by direct measurement at the original research stations and extrapolation procedures were proposed for areas without complete data. These are described in detail in *Agricultural Handbook 282* (Wischmeier and Smith 1965). This includes a nomograph for estimating soil erodibility (K) from soil texture, structure and organic content, graphical solutions for combined slope length and gradient (SL), and values for 128 crop combinations and cropping practices (C). The C factor can be subdivided for variations in cover protection through the cropping cycle. Finally, values of the conservation factor (P) are provided, based on tests at the experimental stations comparing techniques such as contour ploughing or terracing.

The USLE was designed as a conservation guide for farmers to estimate erosion hazard, to identify the most significant contributing factors and to predict the potential reduction in soil loss from introduction of conservation practices. It has been used widely in the United States, and has been effective in the area where original data are available. Subsequent modifications have been incorporated as understanding of soil erosion processes has increased, including, for example, a seasonally adjusted K factor, to reflect changes in soil structure caused by rainfall and weathering.

The modified equation was published in 1991 as the R (Revised) USLE.

The USLE has been useful for soil conservation in the central USA, and as an aid in instruction about the factors which control soil erosion. Unfortunately, its conceptual simplicity has encouraged use in areas for which it was not designed, such as steeply sloping forested lands, or in areas where inadequate rainfall records are available. It is purely empirical, with no proven physical foundation for extrapolation, and it can lead to extremely inaccurate predictions. Attempts have been made to accumulate appropriate data for wider use particularly in the tropics. Where measured data are available it can be used with some confidence, but elsewhere it is not reliable, particularly as a basis for expensive, socially disruptive or contentious measures. It has also been criticized on theoretical grounds as understanding of soil erosion processes has increased. The K factor, for example, is very simplistic, with no attention to important processes such as surface sealing, or to the effect of soil chemistry on erosion resistance. It certainly cannot be used with confidence to predict important soil erosion effects such as contaminant transport, or nutrient enrichment in lakes or streams. Much research has been directed to development of a sound physically based erosion equation, e.g. WEPP (Water Erosion Prediction Project) in the United States and EUROSEM (European Soil Erosion Model) in Europe. Although promising, these are conceptually complex, require data which are often unavailable, and are not yet sufficiently reliable for widespread use.

References

Roose, E.J. (1977) Application of the Universal Soil Loss Equation of Wischmeier and Smith in West Africa, in D.J. Greenland and R. Lal (eds) *Soil Conservation and Management in the Humid Tropics*, 177–187, London: Wiley.

Wischmeier, W.H. and Smith, D.D. (1965) Predicting rainfall-soil erosion losses from cropland east of the Rocky Mountains, *Agricultural Handbook 282*, Agricultural Research Service, United States Department of Agriculture.

Wischmeier, W.H., Smith, D.D. and Uhland, R.E. (1958) Evaluation of factors in the soil-loss equation, *Agricultural Engineering* 39, 458–462, 474.

Further reading

Hudson, N.W. (1981) *Soil Conservation*, London: Batsford.

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

RORKE BRYAN

UNLOADING

'The removal by denudation of overlying material' (Yatsu 1988: 140) was suggested by Gilbert (1904) as a mechanism that produced exfoliation domes in the granites of the Sierra Nevadas, western USA. According to this theory, granites are buried under a deep cover of older rock and so are subjected to compressive stress. That compressive strength was balanced by internal expansive stress competent to cause actual expansion if the external pressure was removed by denudation. This in turn may produce sheeting (EXFOLIATION) that is broadly conformable to topography. However, considerable controversy surrounds the issue of sheeting – does the sheeting produce the topography or is it *vice versa* as the unloading model suggests.

In addition to unloading being a potential cause of joint development, the term unloading is also used to describe the process of weight removal from the crust. *Glacial unloading*, resulting from the wastage of ice caps leads to postglacial faulting and seismicity, and to isostatic compensation. Equally, *erosional* or *denudational unloading* is an important factor in determing uplift and erosion in situations like passive margins (Clift and Lorenzo 1999). *Mechanical unloading*, associated with lithospheric extension, can contribute to flexural uplift of rift flanks (Weissel and Karner 1989).

References

Clift, P.D. and Lorenzo, J.M. (1999) Flexural uploading and uplift along the Côte d'Ivoire–Ghana transform margin, Equatorial Atlantic, *Journal of Geophysical Research* B, 104, 25,257–25,274.

Gilbert, G.K. (1904) Domes and dome structure of the high Sierra, *Geological Society of America Bulletin* 15, 29–36.

Weissel, J.K. and Karner, G.D. (1989) Flexural uplift of rift flanks due to mechanical unloading of the lithosphere during extension, *Journal of Geophysical Research* B, 94, 13,919–13,950.

Yatsu, E. (1988) *The Nature of Weathering*, Tokyo: Sozosha.

A.S. GOUDIE

(URANIUM-THORIUM)/HELIUM ANALYSIS

(Uranium-thorium)/helium analysis in apatite is currently the lowest temperature thermochronometer for providing detailed information on the thermal history of the crust for low (shallow) crustal temperatures. The technique is based on the alpha decay of U and Th in apatite to yield the daughter ^{4}He. (U-Th)/He analysis was the first radiometric dating system developed, by Ernest Rutherford at the beginning of the twentieth century. Independent geological evidence indicated that the ages it returned in its early applications were generally too young, and so the technique was abandoned as a dating tool. It was realized in the 1980s that (U-Th)/He analysis generally returned ages that are too young because the daughter ^{4}He diffuses out of the apatite above c.80 °C, which is well below the 'formation' temperatures that analysts were attempting to date in the technique's early applications. This behaviour means, however, that the (U-Th)/He system can be used to date rock cooling below c.80 °C.

The analytical procedures for (U-Th)/He in apatite are relatively straightforward, involving heating the apatite grains to drive off the daughter ^{4}He for measurement, followed by dissolution of the grains to measure the amounts of the parent U and Th using an Inducting Coupled Plasma Mass Spectrometer (ICPMS). The standard age equation for radioactive decay of parent element(s) to daughter element is then applied. Sample preparation is also relatively straightforward via standard mineral separation procedures, except for the final stage involving microscope handpicking of apatite grains for the analyses. This careful selection of grains is necessary to avoid the analysis of apatites with inclusions of U-bearing minerals, such as zircon. These U-bearing inclusions generate ^{4}He in the apatite but the inclusions' resistance to dissolution may mean that their corresponding U and Th contents remain unmeasured, thereby confounding the age calculation. A correction to the calculated age must be applied to account for the loss of ^{4}He from the outer edges of the grain by recoil (i.e. complete expulsion of ^{4}He from the grain during the alpha decay). This so-called 'recoil correction' requires that grains being analysed be of standard shape and of known size, which is a further reason for careful handpicking and characterization of the crystals to be analysed.

In the same way as there is a temperature range in which fission tracks are only partially retained (the partial annealing zone in FISSION TRACK ANALYSIS), the temperature range between c.80 °C and 40 °C defines the ^{4}He partial retention zone (PRZ). All of the daughter ^{4}He diffuses out of the apatite grain above c.75 °C, but at temperatures cooler than this the He is increasingly retained. Partial retention is grain-size dependent, providing the potential in (U-Th)/He analysis for an analytical tool comparable to the track-length distribution in fission track analysis. The potential of this grain-size effect is still to be fully explored.

Geomorphological applications of (U-Th)/He analysis largely focus on the rates of denudation required to bring apatite-bearing rocks to the Earth's surface from the crustal depths corresponding to a temperature of c.80 °C, at which the ^{4}He daughter product of alpha decay of U and Th begins to be retained. This denudation may be regional continental denudation required to bring the apatite to the Earth's surface (with the only 'tectonic' component being passive denudational isostatic rebound) or the denudation may be driven by tectonic rock uplift. In the latter case, and on the assumption that cooling through the PRZ is coeval with the tectonic uplift, the (U-Th)/He age corresponds to the age of onset of tectonic uplift. The assumption that denudation is coeval with uplift is extremely important. The assumption is reasonable in settings in which agents of subaerial incision, such as fluvial and glacial processes, are efficiently connected to the 'external' reference plane for tectonic uplift (e.g. a local base level or global sea level), and uplift-driven disequilibrium in the drainage net, due to relative lowering of base level, is rapidly transmitted through the drainage net to trigger incision and denudation throughout the catchment. Not all high elevation uplifting areas are well connected to the external base level which is the reference plane for surface uplift (e.g. the Tibet plateau, the Andes Altiplano). In these cases, the low-temperature thermochronological ages of rocks now at the surface of these landscapes may bear little relationship to the onset of uplift.

A further geomorphological application of low-temperature thermochronology, especially (U-Th)/He thermochronology in apatite, relies on the fact that the thermal structure of the shallow crust (upper few kilometres) is deformed by the long-wavelength Earth surface topography. The shallow crustal isotherms mirror topography of length scales of tens of kilometres. Thus the thermal structure of the crust beneath long-lived major valleys, for example, mirrors these valleys. If (U-Th)/He ages along a transect at constant elevation across these valleys (at a mountain front, for example) mimic the topography then the topography must have existed when the apatites were passing through the PRZ. That is, if the (U-Th)/He ages are older on that part of the transect coinciding with the interfluves and younger where the transect coincides with the valleys, the (U-Th)/He ages effectively provide minimum ages for the long wavelength topography.

PAUL BISHOP

URBAN GEOMORPHOLOGY

Urban geomorphology examines the geomorphic constraints on urban development (Cooke 1984) and the suitability of different landforms for specific urban uses; the impacts of urban activities on Earth surface processes, especially during construction; the landforms created by urbanization, including land reclamation and waste disposal; and the geomorphic consequences of the extractive industries in and around urban areas (McCall *et al.* 1996).

Constraints on urban development

The original founders of towns and cities carefully chose their sites for defensive, strategic, resource exploitation, navigation or cultural reasons. Great attention was given to finding sites with adequate water supplies and protection from obvious environmental hazards. However, the growth of these settlements often led to the spread of urban development on to less suitable ground and overstretched the capacity of the local environment to support the community. Many environments have particular conditions that make conditions for modifying slopes or establishing foundations difficult (Table 46).

Application of geomorphological mapping to the classification of the suitability of land for different types of urban development is now part of the work of geological and soil surveys in many countries. Such mapping considers the steepness of slopes, their colluvial and weathered mantles, their drainage and depth to bedrock and provides guidance on the type of development suited to different parts of the slope.

Knowledge of landform evolution is particularly important, as modern earthmoving can reactivate features inherited from past conditions, such as fossil periglacial landslides in Europe and North America. Loading of peat with urban structures formed after the retreat of ice sheets can result in significant subsidence and building damage. Karstic features formed when sea levels were lower in the Quaternary, but now buried under alluvium, can pose severe problems for the foundations of high-rise buildings.

Many present-day conditions constrain urban development. Widespread soils rich in montmorillonitic clays are subject to 'shrink–swell' cracking clay phenomena which require special foundations if buildings are not to become unstable. Climate change is likely to shift the areas where these problems are severe, if summers become drier. Mobile dune systems and sources of wind-blown sand pose problems for the siting of many structures. ALLUVIAL FANs may normally be inactive with the local stream confined to a narrow channel, but they may be reactivated, flooded and covered with debris if an extreme flood descends from the adjacent mountains. Building in PERMAFROST areas has to isolate the heated structures from the frozen ground and be careful not to disturb the permafrost during the construction process.

Geomorphic impacts during urban construction

Urban construction involves removal of the natural vegetation cover and excavation of the topsoil and often much of the underlying weathered rock and bedrock layers. In new urban developments, small streams are often diverted into culverts or urban drains and minor depressions and valleys are filled in. Steep hillsides may be terraced into a series of home sites by cut-and-fill operations. Major rivers may be embanked and artificially straightened. In extreme cases, as in Palma de Mallorca, Spain and Winnipeg, Canada, large new flood channels may be built around the urban centre to divert flows away from the city. The new features replacing the original landforms are often designed to direct water away from the new developments more effectively, so producing off-site, downstream consequences.

The earthmoving operations during urban construction frequently lead to severe erosion problems and consequent channel modifications (Table 47, Figure 173). Erosion control guidelines suggest that construction should be carried out in phases to avoid disturbing too much of the land at any one time. No unnecessary clearing should be undertaken. Immediately below any cleared area, detention ponds should be constructed to retain any sediment washed off the site and to hold back stormwater runoff so that peak discharges in streams below are not increased.

Increased sediment loads and peak storm discharges lead to channel modifications (Figure 173) with formerly meandering streams becoming braided, steeper and shallower. Sometimes these changes are controlled by modifying the channel, often with expensive structural works. However, even these are not always successful as siltation of the channel can follow, with large accumulations of weeds and silt building up if the stream receives discharges with high nutrient loadings. Further downstream, rivers may adjust in response to upstream channelization, eroding their banks, developing new gravel bars and threatening bridge abutments and riverine structures.

Landforms of extraction

Meeting the demand for construction materials changes the land surface, by creating pits and

Table 46 Geomorphological problems for urban development

Environment	Chief problems
A Climatic	
Periglacial	Permanently frozen ground and overlying active layer require special types of construction and foundations for buildings and infrastructure
Arid	Water supply problems; wind erosion; flash floods; possibility of salt weathering of building materials and foundations
Humid tropical	Rapid weathering and decomposition of building materials; deep, uneven weathering of most rocks in tectonically stable areas; frequent rain events causing rapid water erosion of exposed ground surfaces
B Topographic	
Mountainous	Risk of unstable slopes, rockfalls, debris flows and avalanches; potential for flash floods
Floodplains	Liable to periodic flooding; variable foundation conditions over former, buried river channels and alluvial deposits
Coastal plains	Storm surge and flooding risk likely to increase with rising sea levels; complex ground conditions reflecting former shorelines and old drainage channels; possible salt penetration in ground water affecting foundations
Coasts with weak rock cliffs	Liable to rapid coastal erosion, cliff undercutting and collapse; eroded debris often deposited in ports and harbours causing dredging expenditure
Islands	Particular storm-surge, rising sea level and salt water penetration risks on low-lying atolls and coastal plains
C Tectonic/lithological	
Active plate margins	Major risks associated with coastal urban developments, especially on Pacific rim, special foundation requirements on filled areas, lake sediments and other unconsolidated materials; major earthquake-triggered landslide hazards; volcanic debris and lahar risks requiring awareness of flow pathways on lower volcanic slopes likely to have urban settlements
Shrink-swell clays	Cracking clay problems likely to be accentuated by climate change
Karst	Buried karst a major problem for foundations of tall buildings and for sinkhole development; need for knowledge of buried karst plains and effects of lowered Quaternary sea levels

Sources: Based on data in Marker (1996); McCall *et al.* (1996); and Bennett and Doyle (1997)

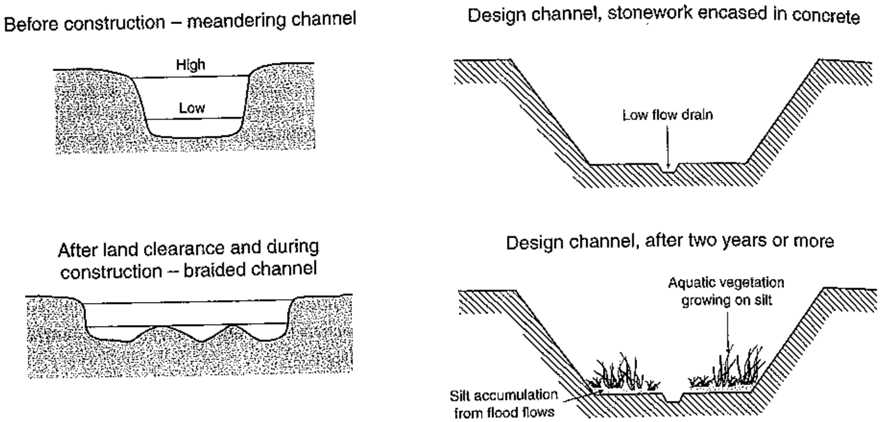


Figure 173 Channel changes to the Sungai Anak Ayer Batu at Jalan Damansara between 1960 and 1990 due to urban development

Table 47 Sequence of fluvial geomorphic response to land use change: Sungai Anak Ayer Batu, Kuala Lumpur

Land cover/land use	Channel condition
Forest	Narrow, meandering with low sediment load
Rubber plantation	Gullying during clear weeding; peak discharge increased; channel slightly widened; later stabilized; few cut-offs
Urban construction	High sediment yield; high peak discharge; metamorphosis to wider, steeper, shallower braided channel
Channelization and stable urban built-up area	Higher peak discharge; less sediment load; channel enlargement downstream; bank erosion, minor channel incision; loss of fine bed material by scour
Post channelization siltation	Where large quantities or organic debris enter concrete channels and are deposited, vegetation can become established and build up deposits that reduce channel capacity

quarries. The largest excavations take up many square kilometres of the land surface, often creating areas of brick pits that sometimes are used for waste disposal, or gravel pits that become peri-urban wetlands, often serving combined recreational and flood control purposes. Not all the filling of former opencut mines is without incident. In the past, escaping methane gas from landfills in old opencast coal pits has caused problems for the houses built upon them. In karstic terrain, the SUBSIDENCE of filled material in chalk pits or in old tin mines overlying cavernous limestone has led to severe damage to houses and urban infrastructure.

Removal of mineral resources and water from beneath the ground leads to subsidence creating new surface topographies and, often, new water bodies. Groundwater pumping has put the historic world heritage buildings of Venice at risk. Built at sea level on the lagoon, Venice has subsided some 22 cm since 1900. Most of that surface lowering occurred between 1950 and 1970. High water (*aqua alta*) has occurred more frequently since 1970. Whilst the people-induced subsidence is part of the problem, higher extreme sea levels related to global climate change may possibly be another factor. In the Los Angeles area, extraction of oil beneath Long Beach created severe subsidence that had to be halted by the injection of water into abandoned wells.

Landforms of deposition

Much modern urban development involves land reclamation and major landform modification (Gupta and Ahmad 2000). In extreme cases, huge quantities of material are moved, for example in the development of major airport sites such as Kansai, Singapore and Hong Kong. At Kansai, the fill material has caused some subsidence of the original seabed, with allowance having to be made for this in the operation and maintenance of the airport. The problems of subsidence of the second-stage runway are expected to be more severe than in the first runway, with a prediction that after fifty years subsidence will have been 18 m compared to 11 m for the first stage. Detailed analyses have been made of the way landing aircraft cause small temporary depressions in the runway that in turn affect the drag on aircraft moving along the runway.

As disposal of solid waste moves from landfill to land raise, new hills appear on the edges of floodplains, above former gravel pits and quarries and on offshore islands. In some urban areas, waste dumps are prominent features of the landscape. Although the older dumps are the result of coal, slate and china clay production, modern land raise features dominate many low relief areas. Whilst much of this waste is deposited in disused opencast mines and quarries, land raise mounds are probably the fastest growing artificial landforms in many countries today. The greatest geomorphological impact of landfill is in river valleys, sections of which are being filled, raising the height of the ground surface well above the former floodplain level. This effectively reduces the flood storage capacity of the floodplain, shifting the flood problems downstream.

Many of the old dumps are being closed or modified, from the huge dumps on the edges of cities like Istanbul and Manila to the managed disposal areas, such as Freshkills on Staten Island, New York, which has been taking nearly all the 17,000 tons of waste the city collects each day. As events at the Payatas tip in Manila have shown, some of these urban waste mounds are unstable, prone to massive slumps and landslides. The loss of life and property that ensues is a challenge to the management of the waste disposal and the application of geomorphology to the construction of land raise mounds.

Urban regeneration itself involves creating new landforms as the old buildings are demolished and construction and demolition waste is used for on-site fill or is taken short distances to sites that have to be raised to be above known flood levels. Many historic city centres have been so rebuilt that the average level of streets is now above that of the entrances to medieval buildings. These changes in landform may often be individually small, but collectively they are the result of two of the main human drivers of global environmental change: increasing urban development and the mining and quarrying which supplies minerals for industry and the construction materials required to build all the infrastructure, homes, offices and factories of cities. Urban geomorphology is thus a key element in supplying the guidance needed to achieve a better quality of urban life and working towards more sustainable use of resources.

References

Cooke, R.U. (1984) *Geomorphological Hazards in Los Angeles*, London: George Allen and Unwin.
Bennett, M.R. and Doyle, P. (1997) *Environmental Geology: Geology and the Human Environment*, Chichester: Wiley.
Gupta, A. and Ahmad, R. (2000) Geomorphology and the urban tropics: building an interface between research and usage, *Geomorphology* 31, 133–149.
McCall, G.J.H., De Mulder, E.F.J. and Marker, B.R. (eds) (1996) *Urban Geoscience*, Rotterdam: Balkema.
Marker, B.R. (1996) The role of the earth sciences in addressing urban resources and constraints, in G.J.H. McCall, E.F.J. De Mulder and B.R. Marker (eds) *Urban Geoscience*, 163–179, Rotterdam: Balkema.

Further reading

Coates, D.R. (ed.) (1976) *Urban Geomorphology*, Geological Society of America Special Paper 174.
Cooke, R.U., Brunsden, D., Doornkamp, J.C. and Jones, D.K.C. (1982) *Urban Geomorphology in Drylands*, Oxford: Oxford University Press.

Douglas, I. (1983) *The Urban Environment*, London: Arnold.
Leggett, R.F. (1973) *Cities and Geology*, New York: McGraw-Hill.

IAN DOUGLAS

URSTROMTÄLER

Broad (2–25 km wide), pronounced depressions which are aligned parallel with the margins of the Pleistocene ice sheets in Europe and North America; also called *ice marginal valleys*, *ice-marginal streamways*. They were first studied in North Germany (Wolsted 1950). These forms can be traced across the North European Plain from Russia to the North Sea. Marshy, flat-floored trenches usually have steep erosive external scarps, 20–40 m high and, often, systems of lateral terraces (Galon 1961). Their courses relate to particular stages of the Scandinavian ice sheet limits. Their origin is usually explained as the product of both erosion and sedimentation by the conjoined waters' from proglacial *meltwater channels* (see MELTWATER AND MELTWATER CHANNEL) and rivers which drained the ice-free areas in the south. Almost certainly, large volumes of water must be involved in their production; at least some might be derived from catastrophic outbursts from ice-dammed lakes. Jahn (1975) considered that intense thermal erosion of riverbanks under permafrost conditions was also a generic factor.

References

Galon, R. (1961) *Morphology of the Notec-Warta (or Torun-Eberswalde) Ice Marginal Streamway*, Warszawa: Prace Geograficzne PAN, No. 29.
Jahn, A. (1975) *Problems of the Periglacial Zone*, Warszawa: PWN-Polish Scientific Publishers.
Woldsted, P. (1950) *Norddeutschland und angrenzende Gebiete im Eiszeitalter*, Stuttgart: Koehler Verlag.

JACEK JANIA

V

VALLEY

'A depression sloping in one direction over all its length' (Von Engeln 1942: 7), which tends to be longer than it is wide. Valleys have a range of sizes and a multiplicity of names – gully, draw, defile, ravine, gulch, hollow, run, arroyo, gorge, canyon, dell, glen, dale and vale (Huggett 2002: 193).

Valleys are normally regarded as the products of a range of fluvial processes such as corrasion, abrasion, pot-holing, cavitation, corrosion and weathering. They widen by lateral stream erosion and by weathering, mass movements and fluvial processes on their sides. They lengthen by such processes as HEADWARD EROSION or by building new land (e.g. RIVER DELTAS) at their bottom ends. In planform they develop networks (see HORTON'S LAWS) and they have a variety of DRAINAGE PATTERNS, including ALIGNED DRAINAGE. Some valleys develop in bedrock (see BEDROCK CHANNEL) while others develop in superficial materials, such as alluvium. In general big rivers have big valleys, but there are cases where small, misfit rivers occupy large valleys. This can be due to river capture, which can divert large amounts of water from that valley into another river system. Alternatively, it can be due to major climatic changes that have decreased the flows of water through the valley meander systems (Dury 1997) while some valleys may have no channels in them at all (see DRY VALLEY). Many valleys are accordant to geological structures, while others are discordant as a result of antecedence or superimposition.

There is a huge diversity to valley forms (see ARROYO; BEHEADED VALLEY; BLIND VALLEY; BOX VALLEY; BURIED VALLEY; CANYON; DAMBO; DELL; MEKGACHA; TUNNEL VALLEY; WADI). While most valleys are of subaerial type, there are also SUBMARINE VALLEYS. Some valleys have regular cross sections, while others, for structural or aspect-related microclimatic reasons, display asymmetry. Equally, normal fluvial valleys are often perceived as having a tendency towards V-shaped cross profiles (though this is far from universal), while glacially excavated valleys are often perceived as giving U-shaped cross profiles with truncated spurs.

The origin of valleys has proved troublesome. In the early nineteenth century they were sometimes regarded as the result of the Noachian deluge (see DILUVUALISM). There was also a great deal of debate as to what extent they were essentially tectonic features, related to fracturing of the Earth's crust. It was not easily recognized that they were the result of rain and rivers. These debates are well reviewed by Chorley *et al.* (1964). However, as Kennedy (1997: 67) has pointed out,

> *Any* process which creates topographic irregularities will cause the subsequent concentration of any available surface moisture and, potentially, a stream-and-valley. Moreover, since valleys are exceptionally durable features... we must face the fact that many networks will contain portions which owe both their ultimate origin and also their persistence to *different* processes.

References

Chorley, R.J., Dunn, A.J. and Beckinsale, R.P. (1964) *The History of the Study of Landforms or the Development of Geomorphology*, Vol. 1, London: Methuen.
Dury, G.H. (1997) The underfit meander problem. Loose ends, in D.R. Stoddart (ed.) *Process and Form in Geomorphology*, 46–59, London: Routledge.
Huggett, R.J. (2002) *Fundamentals of Geomorphology*, London: Routledge.

Kennedy, B.A. (1997) The trouble with valleys, in D.R. Stoddart (ed.) *Process and Form in Geomorphology*, 60–73, London: Routledge.
Von Engeln, O.D. (1942) *Geomorphology: Systematic and Regional*, New York: Macmillan.

A.S. GOUDIE

VALLEY MEANDER

Meanders which are usually cut in bedrock and usually have a greater wavelength than that of the contemporary river pattern. These meanders shape valleys winding rather symmetrically between hills, and they are much wider than the meanders of the river flowing in the alluvial plain or alluvial meanders. The two types of meanders tend to be geometrically similar, the only real difference being that meanders in bedrock are commonly ingrown, whereas meanders on a floodplain are not. Entrenched meanders, which have cut vertically down without enlarging themselves in the lateral and axial directions, are uncommon. They represent one end of a series which extends through the intermediate range of normally ingrown meanders to meanders on a floodplain.

Davis (1906) described bedrock meanders in relation to lateral corrasion and placed their origin during the youth stage of the cycle of erosion. Dury (1954, 1977) thoroughly studied meandering valleys in order to ascertain whether they were cut by a stream larger than the present-day one. From many examples, he showed that valley meanders were produced during periods of higher runoff and higher peak discharges, that is, before stream shrinkage which led to contemporary UNDERFIT STREAMS, and that these larger discharges in past times were associated with Pleistocene climate.

The relation of meander length to valley width in valley meanders in rock shows more scatter compared to those in alluvium, but it is apparent that the length is directly proportional to the channel width in both cases. In the rock meanders wavelength is 15 to 20 times valley width. On the other hand, study of individual bends of valley meanders suggests that differences in geologic structure and lithology lead to differences in wavelength of meanders in rock.

Because of the difficulty of visualizing how a channel could maintain a regular sinuous pattern while cutting across hard-rock strata, it has often been assumed that the meandering pattern was initiated in an overlying sedimentary cover and superimposed on the tougher rock below as the river entrenched itself into the strata. Very often these presumed overlying strata have been eroded away, and hence the hypothesis is difficult to verify. In most cases there appears to be no need for such a two-cycle hypothesis. Other meanders in bedrock suggest that the river was antecedent to uplift. That is, the river appears to have maintained its course, trenching the structure as the latter formed. In the absence of stratigraphic evidence it is impossible to distinguish between an antecedent river and one which was superimposed from an overlying cover.

References

Davis, W.M. (1906) Incised meandering valleys, *Bulletin of the Geological Society of Philadelphia* 4, 182–192.
Dury, G.H. (1954) Contribution to a general theory of meandering valleys, *American Journal of Science*, 252, 193–224.
——(1977) Underfit streams: retrospect, perspect and prospect, in K.J. Gregory (ed). *River Channel Changes*, 281–293, Chichester: Wiley.

Further reading

Leopold, L.B., Wolman, M.A. and Miller, J.P. (1964) *Fluvial Processes in Geomorphology*, San Francisco: W. H. Freeman.

SEE ALSO: underfit stream

MARIA SALA

VASQUE

Limestone and AEOLIANITE SHORE PLATFORMS in Mediterranean and tropical regions commonly consist of a series of low terraces formed by wide, flat-bottomed pools or vasques. The pools are separated from each other by narrow, winding ridges that can be: built by calcareous algae, Vermetids (see VERMETID REEF AND BOILER), or even Serpulids (see SERPULID REEF); the residual CORROSIONal pinnacles of lapiés; or a combination of the two. Plates-formes à vasques are covered at high tide and washed by breaking waves at low tide, with the return flow cascading into successively lower pools (Plate 143).

Plate 143 Rimmed pools, vasques, developed in eroded aeolianites at Treasure Beach near Durban, South Africa

Further reading

Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

VENTIFACT

A term introduced by Evans (1911) to describe wind-faceted stones, the surfaces of which are flattened such that they intersect at sharp angles. They include the brazil-nut shaped 'Dreikanter' of German workers. For their formation three conditions are required: strong, generally unidirectional winds; the presence of loose materials (sand, dust, snow, etc.) that are available for transport in suspension or saltation; and the presence of pebbles, boulders and bedrock outcrops projecting into the wind stream. However, there has been considerable debate as to the relative importance of abrasion by dust and by sand (see Breed *et al.* 1997 for a review) and to the precise mechanisms that produce flattened surfaces on three or more sides of many ventifacts.

Ventifacts have been noted on a wide range of lithologies, including basalt, granite, dolerite, aplite, andesite, chert, marble, dolomite and limestone. They occur in a range of exposed environments, including deserts, periglacial and coastal settings. They also occur on Mars (Bridges *et al.* 1999). Some ventifacts are relicts of former tundra conditions subsequent to ice recession but before vegetation establishment in the Late Pleistocene. Such ventifacts have been used to infer palaeowind directions (Schlyter 1995), with strong easterly winds being present in Denmark and southern Sweden. In Ireland, some coastal ventifacts may have formed during the Little Ice Age of the Late Holocene, when there were increased offshore winds, waves, sediment fluxes and periods of sand dune construction (Knight and Burningham 2001).

References

Breed, C.S., McCauley, J.F., Whitney, M.F., Tchakerian, V.P. and Laity, J.E. (1997) Wind erosion in drylands, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology: Process, Form and Change in Drylands*, 437–464, Chichester: Wiley.
Bridges, N.T., Greeley, R., Haldemann, A.F.C., Herkenhoff, K.E., Kraft, M., Parker, T.J. and Ward, A.W. (1999) Ventifacts at the Pathfinder landing site, *Journal of Geophysical Research* 104(E), 8,595–8,615.
Evans, J.W. (1911) Dreikanter. *Geological Magazine* 8, 334–345.
Knight, J. and Burningham, H. (2001) Formation of bedrock-cut ventifacts and Late Holocene coastal zone evolution, County Donegal, Ireland, *Journal of Geology* 109, 647–660.
Schlyter, P. (1995) Ventifacts as palaeo-wind indicators in southern Scandinavia, *Permafrost and Periglacial Processes* 6, 207–219.

A.S. GOUDIE

VERMETID REEF AND BOILER

CORNICHES and other organic REEFs in the northwestern Mediterranean usually consist of calcareous algae and Serpulid (see SERPULID REEF) worms. Higher temperatures in the southern Mediterranean are favourable for large Vermetid populations, but whereas they only form veneers on TROTTOIRs in fairly easily eroded limestones and sandstones, there are purely constructional Vermetid corniches on fairly resistant substrates. Vermetids also contribute to the development of boilers or cup reefs in the Mediterranean and western Atlantic. Boilers, which are awash at low tide, are up to 12 m in height and a few tens of metres in diameter. They resemble MICROATOLLs with a central depression or micro-lagoon, up to a few metres in depth, surrounded by raised rims. Boilers in Bermuda consist entirely of coralline algae, Vermetid gastropods and the encrusting coral *Millepora*, but similar forms in the Mediterranean are merely veneers of Vermetids and algae over eroded AEOLIANITE blocks.

Further reading

Ginsburg, R.N. and Schroeder, J.H. (1973) Growth and submarine fossilization of algal cup reefs, Bermuda, *Sedimentology* 20, 575–614.
Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.

ALAN TRENHAILE

VISOR, PLINTH AND GUTTER

CORROSIONal notches (see NOTCH, COASTAL) at the cliff foot sometimes have protruding visors above them and plinths below them. They have been described on AEOLIANITE SHORE PLATFORMS in southern Australia (Hills 1971), and from western Australia, Hawaii, Bermuda and northwestern India. The visor consists of a band of hardened, indurated rock, which may form when fresh rain water deposits calcium carbonate where it comes into contact with rock that is saturated with sea water. This might explain why the height of the visor declines as it is traced into sheltered areas, although it is questionable whether seawater saturation in the high and supratidal zones can be maintained during low tidal periods. The plinth is a slight prominence which is attached to the outer edge of the notch base. Hills proposed that the plinth develops at the height to which water is drawn by capillary action above the platform surface. The gutter, or moat, is a channel, occasionally found at the base of a ramp (see RAMP, COASTAL), that has been eroded by sand, pebbles and small boulders.

Further reading

Hills, E.S. (1971) A study of cliffy coastal profiles based on examples in Victoria, Australia, *Zeitschrift für Geomorphologie* 15, 137–180.

ALAN TRENHAILE

VOLCANIC KARST

Karst-like features occur in volcanic terrains and they have been classified into four types (Reffay 2001). These are shown in Table 48.

The type *pseudokarst sensu lato* consists of structural landforms in lava flows that are unrelated to denudational processes. They include lava tubes, lava speleothems, and pseudodolines and shafts generated by collapse of lava flow roofs. The type *pseudokarst sensu stricto* is created by piping processes in loose volcanic material (e.g. pyroclastic deposits). The type *orthokarst* develops as a result of dissolution of carbonatites. The type *parakarst*, develops as a result of dissolution of minerals other than carbonates.

Reference

Reffay, A. (2001) Types de karst en terrain volcanique: revue bibliographique, *Géomorphologie* 2001(2), 121–126.

A.S. GOUDIE

Table 48 Volcanic karst classification

Pseudokarst, S.L.	Lava	Tunnels, speleothems, etc.	Structural. Related to lava emplacement
Pseudokarst; S.S.	Pyroclastics; Andosols	Pipes, holes, canyons, dry valleys	Suffosion
Orthokarst	Carbonatites	Closed depressions, megalapiés	Dissolution of carbonates
Parakarst	Basalt	Closed depressions, lapiés, speleothems, travertine and sinter	Dissolution of Ca, Mg, Na and silica. Precipitation of $CaCO_3$ and SiO_2

Source: Modified from Reffay (2001)

VOLCANO

A volcano can be defined as the site on the surface of a planet or moon through which gaseous, liquid and/or solid materials are expelled or erupted, usually through the action of internal thermal processes. Eruptions often discharge magma – molten igneous material composed of a silicate or other liquid, with variable quantities of crystalline phases and gas bubbles, though this can be transformed dramatically in violent, explosive eruptions into a stream of rock fragments and hot gases. An eruption can also result from steam explosions that blast out near-surface

rocks without any accompanying fresh magma. The erupted products typically accumulate around the eruptive vent or vents, and can, if eruptions are sustained or repeated, construct mountains of very considerable volume. Olympus Mons, the highest volcano in the Solar System, rises some 24 km above the surrounding martian plains, and has a volume of around 3×10^6 km^3 (Plate 144). Volcanism, past and present, is one of the fundamental geological processes of the Solar System.

On Earth, volcanoes are broadly distinguished as either active or extinct. The term 'active' is used in different senses. Often it is used to indicate a volcano actually in eruption. However, it is also applied to all volcanoes known to have erupted in the Holocene period (last 10,000 yr). This broader definition obviously includes many volcanoes that have not erupted for centuries or even millennia, and which may actually be extinct (incapable of future eruption) but it helpfully covers many more volcanoes, which may experience long repose periods (the intervals between eruptions), and which can be considered dormant and capable of further eruption. Around 1,500 volcanoes are known or suspected to have erupted during the Holocene,

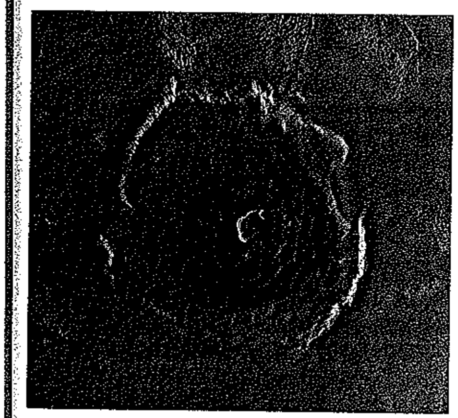


Plate 144 Olympus Mons – highest volcano in the Solar System. Note the overall shield shape, broad summit region crowned with nested calderas, and the prominent 550 km circumference, several km high basal scarp, whose origin has been the subject of much debate. Image processed by J. Swann, T. Becker and A. McEwen, and archived by NASA/NSSDC

and are therefore classed as active (Siebert and Simkin 2002). Of these, some 550 erupted in the historic period. Every year, an average of fifty to seventy volcanoes erupt, though some of these are volcanoes with long-lived eruptions spanning years or decades. On an average day, at least twenty of Earth's volcanoes will be erupting as you read this page. It is important to point out that all these figures are for volcanoes on land. Accurate figures for submarine eruptions are not available, although it is known that seafloor volcanism dwarfs the magma output of subaerial volcanism by a ratio of as much as ten-to-one.

Many dormant volcanoes discharge gases and liquids at the surface. In the case of hot springs, the flux of steam and gas is subordinate to that of liquid water. Geysers are spectacular examples of hot springs, Old Faithful in Yellowstone, Wyoming (USA) being perhaps the most famous. When gaseous emissions predominate, the term fumarole is usually applied. Emission temperatures of fumaroles therefore typically exceed the local boiling point of water. Long-lived fumarole fields are sometimes termed solfataras or soufrieres. Fumarole emissions are very often composed of both magmatic and hydrothermal gases, the latter evolving through complex chemical and physical interactions between magmatic fluids, meteoric water, seawater and rock. The emissions of some volcanoes discharge into crater lakes, which form by the condensation of the gases in the lake, as well as the capture of precipitation. The 'black smokers' associated with oceanic ridges are another important manifestation of subaqueous volatile discharge.

Some volcanoes have been in apparently continuous eruption for as long as records exist. For example, there is no evidence for any significant hiatus in the ongoing eruption of Stromboli (Italy) in over two millennia. The following volcanoes have been erupting more or less continuously for decades: Stromboli and Etna (Italy); Erta 'Ale (Ethiopia); Manam, Langila and Bagana (Papua New Guinea); Yasur (Vanuatu); Semeru and Dukono (Indonesia); Sakura-jima (Japan); Santa Maria and Pacaya (Guatemala); Arenal (Costa Rica); Sangay (Ecuador); and Erebus (Antarctica). According to the records of the Smithsonian Institution's Global Volcanism Program, the median duration of an eruption is around seven weeks. Most eruptions end within three months.

Materials erupted

Excluding some exotic but rare instances, most volcanoes erupt silicate magma of one sort or another. On eruption, these materials can be divided into lavas, which flow (with widely varying degrees of ease) across the surface in partially molten form, and pyroclastics (literally 'broken by fire'), which are expelled from volcanoes as solid fragments. Pyroclasts may be lofted in buoyant eruption columns to considerable heights in the Earth's atmosphere (exceeding 40 km altitude in exceptional cases), and can then be transported hundreds or thousands of kilometres by air currents. When they sediment to the surface they typically form ash fall deposits. These are often characterized by superposed beds of well-to-moderately sorted ash. If ash deposits harden, they are often called tuffs.

Although the term 'ash' is used widely in a loose sense, strictly it refers to pyroclasts with a diameter of 2 mm or less. The very finest ash is composed of shards, minute fragments of volcanic glass shattered apart by violent, explosive eruptions. Larger fragments (up to 64 mm across) are termed lapilli, and still larger ejecta are referred to as blocks (if solid when ejected) or bombs (partly molten when ejected). The very largest blocks and bombs tend to separate from an ascending eruption column and follow ballistic trajectories back to the surface, dropping relatively close to the vent. Fluid, coarse pyroclasts that accumulate around the vent are often termed spatter. Lapilli of mafic or intermediate composition (low or moderate silica content) that display a fragmented, bubbly texture are termed scoria. Highly bubbly pyroclasts, usually of more silica-rich (silicic) composition, are often called pumice. A common term pertaining to all pyroclastic material regardless of size is tephra.

Pyroclastic eruptions do not always produce stably convecting eruption columns in the atmosphere. They can also result in fountain collapse, in which all or part of the jet of pyroclasts leaving the vent region is unable to entrain and heat sufficient air to become buoyant. It then sinks back to the surface, where it can feed pyroclastic flows that move across the ground as a density current. Generally pyroclastic flows follow topography closely but they can gain sufficient momentum to overcome topographic barriers. The largest silicic eruptions on Earth are predominantly pyroclastic flow eruptions with volumes of a few thousand cubic kilometres (masses exceeding 10^{15} kg). Their deposits are often termed ignimbrites, and some anneal during compaction to form an excellent source of building stone. Pyroclastic flows consisting of ash-sized material are also known as ash flows, and their sediments as ash flow deposits. The finest particles can separate from a pyroclastic flow, and rise as a convecting plume to considerable heights in the atmosphere, ultimately settling out to form co-ignimbrite ash deposits.

Eruption styles

We have already referred to two very generalized eruption styles – lava (effusive) and pyroclastic (explosive) eruptions. Various schemes exist for classifying eruptions in finer detail. While they all retain some descriptive value, confusion can arise from inconsistent use of the terminology, and the fact that an individual eruption can display many different phenomena in rapid succession or even simultaneously (Plate 145). There are two basic approaches – one to describe an eruption on the basis of contemporaneous observations (i.e. the physical description of the eruptive phenomena), the other to characterize and interpret the deposits or impacts of an eruption. Clearly, the latter is the best or only available option for many eruptions in the past, and considerable efforts have gone into developing theoretical frameworks relating the depositional record to eruption physics and the atmospheric transport of ash clouds (Sparks *et al.* 1997).

Plate 145 Mount Etna (Italy) in eruption in August 2001. Note that several aligned vents are active simultaneously. The one disgorging dark ash clouds is at 2,550 m above sea level and produced a new cinder cone, now a tourist attraction. The vent below it is emitting a lava flow, picked out by the curtain of whitish gases rising above it

The least ambiguous physical descriptors of eruptions are magnitude, intensity and duration. Intensity describes the mass eruption rate (e.g. in kg s^{-1}), and is a particularly useful parameter for explosive eruptions, as it is closely related to the height reached by the eruption column. Integrating intensity through the whole duration of an eruption yields the total erupted mass, or magnitude (in kg). This is also a useful first-order measure that enables intercomparison of sizes of different lava and/or tephra eruptions.

Eruptions do not always expel fresh magma. They can be driven by the sudden expansion of a liquid changing phase into a gas – for example, when ground water comes into contact with magma and flashes to steam – or by the instantaneous decompression of pockets of gas that have accumulated at some position within a volcano or its basement. These phreatic blasts can excavate considerable volumes of rock between the explosion source and the Earth's surface, leaving impressive holes in the ground. If new magma is also expelled in such explosions, they are termed phreatomagmatic. Such hydrovolcanic phenomena are quite common when a dormant volcano awakens – the volcano is effectively clearing its throat to make way for the passage of new juvenile magma.

A suite of more subjective terms to describe eruption style is also in common usage. These derive from particular historic eruptions (for example, 'vulcanian' refers to the 1888–1890 eruption of Vulcano; 'plinian' to Pliny the Elder's observations of the AD 79 eruption of Vesuvio, recorded by his nephew) or characteristic styles of individual volcanoes ('strombolian' refers to Stromboli volcano's propensity for modest pyrotechnic displays). Unfortunately, because one volcanologist's idea of a vulcanian eruption can be rather like another's image of a plinian, the terminology is not ideal but since it remains in widespread use it will be briefly outlined here.

Gas-rich eruptions of low viscosity magma can be sustained, generating the fire fountains characteristic of Hawaiian activity. Strombolian activity is typified by more discrete, fairly instantaneous explosions capable of propelling bombs a few hundred metres above the vent. These eruptions are formed as large bubbles of gases burst at the surface of a conduit filled with magma. Vulcanian eruptions are more violent. Here, build-up of gas pressure, often in a volcanic pipe blocked for decades or centuries since the last eruption, suddenly blows out a dense column of blocks and ashes, often composed of more old lava than new. These sometimes turn out to be throat-clearing eruptions that clean out the volcanic conduit ready for a more substantial plinian eruption. Plinian eruptions typically last for hours or days. The eruption plumes soar to heights of 20 to 40 km, and ash, gases and aerosols can circle the globe in a matter of weeks. Plinian eruptions produce well-sorted pumice and ash fall deposits. One important factor, as explosive eruptions crank up in scale, is the relationship between the duration of magma discharge and the rise time of the plume in the atmosphere. The physics of eruption columns diverge for discrete vs. sustained eruptions, with the latter capable of significantly higher ascent for a given intensity. The most intense known eruption, based on studies of its deposits, is the *c.* AD 181 outburst of Taupo in New Zealand. With an estimated intensity exceeding 10^9 kg s^{-1}, its ash cloud would have climbed to around 50 km above sea level (Carey and Sigurdsson 1989). Occasionally, as at Mount St Helens (USA) in 1980, volcanic explosions are directed more or less horizontally, and are termed lateral blasts.

Hydrovolcanic eruptions are sometimes referred to as surtseyan events. Lava dome eruptions (see LAVA LANDFORM), which often show sudden switches in behaviour between slow effusion of lava that accumulates in the dome, and explosions and dome collapses that feed pyroclastic flows, are termed peléean. Fissure eruptions have yet to earn a special name but they are quite common in some volcanic regions, especially Iceland, where magmas can rise up to the surface in vertical sheets (dykes) of considerable length. The discharge of magma quickly focuses on a number of discrete but aligned points, and the total length of the system can be up to 10 km or more.

Types of volcanoes

The landforms constructed by volcanism are very diverse, reflecting the supply rates of melt from the mantle, the storage, evolution and transport of magma in the crust, and the tectonic environment, as well as external factors such as presence of liquid water. Perhaps the simplest volcanic construct is the cinder or scoria cone. These are often monogenetic volcanoes formed as the result of a single eruptive episode. They seldom exceed 100–200 m in height and are typically composed of mafic scoria. Many much larger shield volcanoes, characterized by low angle, convex-upwards slopes and a broad summit region (Plate 144), are dotted with adventive cinder cones, which develop where branches from the central magma conduit break the surface on the flanks of the volcano. The shield profile reflects the generally low viscosity of the erupted lavas, and their ability to flow for considerable distances before solidifying. Shield volcanoes are usually crowned by nested and intersecting CALDERAS and collapse pits formed by subsidence. Calderas can also develop during large explosive eruptions as the crust founders above the emptying magma chamber.

When there is abundant ground water, hydrovolcanic eruptions can blast out wide depressions enclosed by low, circular or elliptical rims of ejecta. These features are known as maars and they may show little or no juvenile material. Tuff rings are distinguished from maars by being built on to the substrate rather than excavated into it, and typically contain more juvenile tephra. They have gentler slopes (2–10°) compared with tuff cones, which have gradients of 20–30°.

Most volcanoes are polygenetic, the result of numerous eruptive episodes. Simple cones, also called strato-volcanoes, are usually composed of interlayered lavas and tephra produced in countless eruptions over the lifetime of a volcano, which can be anywhere from a few thousand to many hundreds of thousands of years. They typically range in height between 1,000 and 3,000 m, and are often crowned by comparatively small craters (a few hundred metres in diameter). Mount Mayon (Philippines) is a good example, and justly famed for its near-perfect radial symmetry and convex-upwards slopes. Volcanoes may be subjected to major gravitational collapses during their history (see next section) but regrow by further eruption. The remodelled edifices that result are termed composite cones.

Clusters of overlapping volcanoes are sometimes called volcanic complexes, though the term is used rather loosely. Another case of more distributed volcanism is the volcanic field, formed where a single magmatic system has fed many discrete, usually monogenetic eruptions. The Michoacán–Guanajuato volcanic field in Mexico, composed of some 1,400 individual cinder cones, tuff cones and maars peppering a 200 × 250 km area, is a fine example.

Submarine volcanism has only recently received much attention, thanks to advances in deep-sea exploration. The geomorphology of ocean ridge volcanoes displays some of the characteristics of their subaerial counterparts in extensional tectonic environments but eruption in water at high pressure suppresses explosive activity and causes lava surfaces to solidify rapidly. Similar considerations apply to subglacial eruptions, common in Iceland. Here the weight of ice and presence of meltwater result in formation of tuyas in the case of central vent eruptions, or móbergs when fissure eruptions take place. Catastrophic releases of meltwater, jökulhlaups, associated with subglacial eruptions, are responsible for some of the highest discharge rates ever measured, and are capable of transporting massive quantities of sediment. In Iceland the deposits form plains known as sandur.

Erosional features

Because of their physical prominence, volcanoes are prone to all the usual agents of erosion. There can even be feedbacks between constructive and destructive processes that strongly affect the history of a volcano and its geomorphology. Measures of the degree of erosion of a volcano by wind, rain or ice, can be usefully applied to assessment of the relative age of activity. Rivers can slice rapidly through pyroclastic deposits on a volcano's flanks, while fresh tephra can even more quickly infill them. Once erosion gets the upper hand, volcanoes can exhibit well-developed planezes, triangular facets on the flanks of the cone delineated by the intersection of gully heads.

The most devastating destructive events that affect volcanoes are large-scale gravitational collapses, or volcanic landslides. The largest flank failures, sometimes called sector collapses, are a common feature of many composite cones and oceanic volcanoes, and they have generated the largest debris avalanche deposits on Earth (Moore *et al.* 1989). These are often identifiable from their characteristic hummocky topography, even when traces of the avalanche scar have been obliterated.

References

Carey, S. and Sigurdsson, H. (1989) The intensity of plinian eruptions, *Bulletin of Volcanology* 51, 28–40.

Moore, J.G., Clague, D.A., Holcomb, R.T., Lipman, P.W., Normark, W.R. and Torresan, M.E. (1989) Prodigious submarine landslides on the Hawaiian Ridge, *Journal of Geophysical Research* 94, 17,465–17,484.

Siebert, L. and Simkin, T. (2002) *Volcanoes of the World: An Illustrated Catalog of Holocene Volcanoes and their Eruptions*, Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3, (http://www.volcano.si.edu/gvp/world/).

Sparks, R.S.J., Bursik, M.I., Carey, S.N., Gilbert, J.S., Glaze, L., Sigurdsson, H. and Woods, A.W. (1997) *Volcanic Plumes*, New York: Wiley.

Further reading

Francis, P. and Oppenheimer, C. (2004) *Volcanoes*, Oxford: Oxford University Press.

Heiken, G. and Wohletz, K. (1985) *Volcanic Ash*, Berkeley: University of California Press.

Sigurdsson, H., Houghton, B.F., McNutt, S.R., Rymer, H. and Stix, J. (eds) (2000) *Encyclopedia of Volcanoes*, San Diego: Academic Press.

Thouret, J.-C. (1999) Volcanic Geomorphology – an overview, *Earth-Science Reviews* 47, 95–131.

SEE ALSO: caldera; lava landform

CLIVE OPPENHEIMER

W

WADI

A fluvial valley in a dryland region. Some wadis are relict landscape features that developed in former pluvials either as a result of increased overland flow or because of groundwater sapping (see DRY VALLEY). They include the *Megakcha* of the Central Kalahari, an area where there is currently little or no surface runoff. In the Sahara some of the wadis are enormous palaeodrainage systems up to 1,400 km in length (Pachur and Peters 2001). Others are more stubby features that some authors attribute to groundwater sapping (e.g. Luo *et al.* 1997).

Other desert wadis are sporadically active systems. This is, for example, the case in the Negev, where runoff and sediment yields can be high. One reason for this is the nature of rainfall events in the region. Rainfall intensities can be high (Schick 1988). In the Nahel Yael catchment, over a seventeen-year period, intensities exceeding 14 mm hr^{-1} accounted for nearly one-half of the total rain (223 mm out of 449). Of this intense rain, 37 per cent fell in intensities exceeding 2 mm min^{-1}. Extreme flooding in the wadis can follow major rainfall events, as was demonstrated by the storms that afflicted southern Israel and Jordan in 1966 (Schick 1971).

Not all desert rainfall occurs in intense storms. A major contributing factor in runoff generation is the nature of some of the desert surfaces. For example, with dry conditions and a limited vegetation cover, silty soils, associated with loess deposits, rapidly become sealed under the influence of raindrop impact, and so have diminished infiltration capacities. Even on moderate slopes, silty soils generate substantial runoff (Evenari *et al.* 1983). Another runoff generating surface type results from the presence of organic crusts. These contain Cyanobacteria which partially plug soil pore space, particularly when they swell after they are moistened by rain (Verrecchia *et al.* 1995). Yet another important type of surface for runoff generation is bare rock. Available data indicate that the threshold level of daily rainfall necessary to generate runoff in rocky areas is a mere 1–3 mm. This compares with 3–5 mm for stony colluvial soils and more than 10 mm for stoneless loess soils. Because arid areas have a greater exposure of bare rock than semi-arid their wadis may generate more runoff and sediment (Yair and Enzel 1987).

Studies of experimental catchments over several decades have indicated high rates of sediment yield. Of particular note is the magnitude of bedload transport that has been recorded in the Nahal Yatir and neighbouring areas. Reid *et al.* (1998) showed that although wadi channels are only hydrologically active for about 2 per cent of the time (*c.*7 days per year) and only have overbank flow for about 0.03 per cent of the time (3 hours per year), the bedload flux is remarkably high. Indeed, the Nahal Yatir is about 400 times more effective at transporting coarse material than its perennial counterparts in humid zones (Laronne and Reid 1993). The explanation for this (Reid and Laronne 1995) is that its bed is not armoured (see FLUVIAL ARMOUR) with coarse material. The unvegetated nature of the desert watershed provides ample supplies of sediment of all sizes and this, together with the rapid recession of the flash flood hydrographs and the extended periods of no flow, discourages the development of an armour layer. Therefore, the flux rates are not sediment-supply limited as they so often are in perennial stream channels.

References

Evenari, M., Shanan, L. and Tadmor, N.H. (1983) *The Negev: The Challenge of a Desert*, 2nd edition, Cambridge, MA: Harvard University Press.

Laronne, J.B. and Reid, I. (1993) Very high rates of bedload sediment transport by ephemeral desert rivers, *Nature* 366, 148–150.

Luo, W., Arvidson, R.E., Sultan, M., Becker, R., Crombie, M.K., Sturchio, N. and El Affy, Z. (1997) Groundwater-sapping processes, Western Desert, Egypt, *Geological Society of America Bulletin* 109, 43–62.

Pachur, H.-J. and Peters, J. (2001) The position of the Murzuq Sand Sea in the palaeodrainage system of the Eastern Sahara, *Palaeoecology of Africa* 27, 259–290.

Reid, I. and Laronne, J.B. (1995) Bedload sediment transport in an ephemeral stream and a comparison with seasonal and perennial counterparts, *Water Resources Research* 31, 773–781.

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

Schick, A. (1971) A desert flood: physical characteristics; effects on man, geomorphic significance, human adaptation. A case study of the Southern Arava watershed, *Jerusalem Studies in Geography* 2, 91–155.

——(1988) Hydrological aspects of floods in extreme arid environments, in V.R. Baker, R.C. Kochel and P.C. Patton (eds) *Flood Geomorphology*, 189–203, New York: Wiley.

Verrecchia, E., Yair, A., Kidron, G.J. and Verrecchia, K. (1995) Physical properties of the psammophile cryptogamic crust and their consequences to the water regime of sandy soils, north-western Negev Desert, Israel, *Journal of Arid Environments* 29, 427–437.

Yair, A. and Enzel, Y. (1987) The relationship between annual rainfall and sediment yield in arid and semi-arid areas. The case of the Negev, *Catena Supplement* 10, 121–135.

A.S. GOUDIE

WATER-LAYER WEATHERING

Water-layer weathering refers to the accelerated geochemical weathering that occurs on SHORE PLATFORMS immediately above the platform water level. The weathering is a result of a number of interrelated processes that require an unsaturated or alternately wet and dry environment in which to operate. These processes include the combined actions of wetting and drying including thermal expansion in some rocks, the chemical action of salt spray, salt crystallization and the removal of solutions through rock capillaries. Additional processes acting in the same environment can include wave and rock abrasion, biological processes including rock boring, and frost shattering. The main role of waves however is to remove the debris provided by the weathering processes. Over time positive relief features on the platform are weathered down to the water level. Below the water level (the level of saturation) the lack of drying and free oxygen precluded the above processes.

The type and extent of individual processes will be dependent on latitude/climate, exposure to sun or orientation, lithology and rock structure, tide range and level of wave energy on the platform. Water-layer weathering will progress most rapidly on platforms exposed to regular wetting by seawater or spray, in warm temperate to tropical environments which favour drying, and in weaker and more porous sedimentary rocks which increase the depth of activity and aid removal of debris.

Further reading

Stephenson, W.J. and Kirk, R.M. (2000) Development of shore platforms on Kaikoura Peninsula, South Island, New Zealand. II: The role of subaerial weathering, *Geomorphology* 32, 43–45.

SEE ALSO: chemical weathering; shore platform; wetting and drying weathering; weathering

ANDREW D. SHORT

WATERFALL

A waterfall can be distinguished from cascades, cataracts or other sharp descents in a stream profile, by the free fall of water over a very steep rock face. The greatest single fall is 986 m at Angel Falls in Venezuela; Tugela Falls in South Africa drops 948 m; and a descent of 800 m occurs in several drops at Yosemite Falls. Although they have much smaller descents, Victoria Falls (123 m), Niagara Falls (62 m), Iguazu Falls (70 m) on the Parana River and Khone Falls on the Mekong River (22 m) are notable for their great discharges.

Many waterfalls have been initiated by tectonic uplift along continental margins, or by local tectonic disruption along a fault (see FAULT AND FAULT SCARP) or rift scarp, as at Thingvellir in Iceland. They also commonly occur where a stream profile has been steepened by glacial erosion of a valley side, as at Skykje Falls in the Hardanger Fjord of Norway, or where streams flow over sea cliffs. Some small waterfalls, like the upper Suha Falls (25 m) in Slovenia, are essentially constructional features resulting from the deposition of tufa on the bed of a stream. The great majority, however, are the result of the differential erosion of rocks of varying strength.

The example most commonly cited is Niagara Falls, probably because of the fine account written by G.K. Gilbert. There, a dolomite caprock is undercut by failure in the shale beneath it. Other notable examples of caprock waterfalls are Gullfoss and Dettifoss, in Iceland, though in these cases the fall is due to the variable resistance of basalt, and of interbasaltic sediment and breccia. But not all waterfalls are undercut. Many have vertical faces, and others are buttressed outward at the base. These types of falls are not just short-lived features, where a stream is temporarily held up by a resistant vertical barrier. Many of them have retreated considerable distances. Vertical or buttressed forms are widespread in south-east Australia, where they occur in crystalline and in sedimentary rocks; Dangar Falls and Fitzroy Falls are typical examples.

Since Gilbert's account of Niagara, undercutting of waterfalls has been widely attributed to the erosive power of water swirling back into the recessed fall face. However, in most cases this does not occur, because, especially during flood discharge, the water descends well out from the fall face. Most undercuts are quite dry, except when spray is carried in by up-valley winds. Spray may promote weathering of rock in an undercut, especially if it freezes and fractures the rock. Seepage is also important. When saturated, a rock may have only about half its normal strength. Moreover, the abrupt drop in elevation at a large waterfall may greatly increase the pore-water pressure caused by seepage within the rock at the base of the falls. For example, the large undercut at Belmore Falls, in south-east Australia, occurs not behind the falling water, but to one side of it, where the main line of seepage drops 60 m from the valley above.

Water flowing over the fall face can pluck joint-bounded blocks from the rock on the lip of the falls. During flood discharge, when velocities are great, erosion on the lip may also be caused by processes like CAVITATION. The main effect of the falling water is the excavation of a plunge pool at the base of the falls. The claim that falling water has little erosive power and that the erosion is primarily the result of the impact of transported boulders has been repeated many times, but is not true. The kinetic energy of debris-free water falling over a dam can result in serious erosion at the base. For example, water over the Kariba Dam on the Zambezi River scoured a hole 50 m deep in just four years. The maximum, or limiting, depth of erosion varies with the height of the fall and with the magnitude of the discharge, together with the strength of the rock into which the pool is cut. Debris carried over many falls, or scoured from the plunge pools, forms a protective armouring of bed at the downstream end of the pool.

The basic requirement for a waterfall to develop is that the rock incised by a stream has sufficient strength to stand in a steep face. As stress related to the weight of rock is greatest near the bottom of a steep face, a waterfall cut even in an essentially homogeneous rock, such as granite, is likely to fail at its base, and thereby retreat upstream. Even where weaker rocks occur on an undercut waterfall, they commonly are quite fresh, and rupture by brittle fracturing in response to high stress. For example, stress in the gorge walls below Niagara Falls is causing the shale beneath the dolomite to bulge outwards and to generate vertical fractures, which weaken the rock face.

The rates at which waterfalls retreat depend largely on the magnitude of their discharge and the resistance of the rock over which they tumble. It also depends on their planimetric shape, for horizontal stresses normally are such that a waterfall in a narrow slot is less stable than one flowing over a broad AMPHITHEATRE. Rates of retreat range from about 1 km per 1,000 years at Niagara and Victoria Falls, to about 0.1–2 m per 1,000 years at smaller waterfalls in south-east Australia.

References

Gilbert, G.K. (1896) Niagara Falls and their history, in National Geographic Society *The Physiography of the United States*, 203–236, New York: American Book.

Lee, C.F. (1978) Stress relief and cliff stability at a power station near Niagara Falls, *Engineering Geology* 12, 193–204.

Schwarzbach, M. (1967) Islandische Wasserfalle und eine genetische Systematik der Wasserfalle uberhaupt, *Zeitschrift für Geomorphologie* 11, 377–417.

Weissel, J.K. and Seidl, M.A. (1998) Inland propagation of erosional escarpments and river profile evolution across the southeast Australian passive continental margin, in K.J. Tinkler and E.E. Wohl (eds) *Rivers over Rock: Fluvial Processes in Bedrock Channels*, 189–206, Washington: American Geophysical Union.

Young, R.W. (1985) Waterfalls: form and process, *Zeitschrift für Geomorphologie Supplementband* 55, 81–95.

R.W. YOUNG

WATERSHED

The term watershed in British English is used for a drainage divide – watershed boundary (American English) or catchment divide (Australian English) – that defines a water parting or a line, ridge or summit of high ground separating the water flow in two different directions draining in two drainage basins or catchments. In American English and by definitions of several international organizations the term has been changed to signify the region drained by, or contributing water to, a stream, lake or other body of water and is often used synonymously with the term drainage basin or catchment (Bates and Jackson 1980). A watershed can direct water from a surrounding watershed to a point such as a watershed outlet or a sinkhole. The total area of the watershed in a horizontal projection above a discharge-measuring point (watershed outlet) is the watershed area, catchment area or basin area. The watershed can be subdivided into several subwatersheds or subcatchments that conduct water from the surrounding hillslopes from one or the other side into a channel or drain water from a tributary channel at a confluence into a larger channel of a higher order or magnitude (see STREAM ORDERING).

One source of local watershed leakage is seepage or underground flow of water from one drainage basin to an outlet in a neighbouring drainage basin during a flood. Sometimes stream flow in a karst area will flow into an underground channel or sinkhole or directly into the sea. This may also occur when a water body is involved, such as subterranean flow or seepage from a stream, swamp or a lake (drainage lake) into a neighbouring watershed, or above the surface, e.g. stream bifurcation as in deltas or braided river systems (see DRAINAGE PATTERN). In such cases it is very difficult to exactly locate the surface and subsurface watershed boundary. In mountainous regions one therefore defines a watershed that runs along the crest of the highest range as a normal watershed in contrast to an anomalous watershed that does not behave that way. The location of the watershed boundary and size is dynamic over time and space depending on the migration of the watershed boundary caused by, e.g. water body level (overtopping the divide), temperature (thermokarst), sedimentation (alluvial fan), isostatic movements or disruptions by an earthquake or MASS MOVEMENTS, that temporarily or permanently changes the flow direction (Fairbridge 1968).

A watershed can be considered as a dynamic environmental system unit that has the functional organization of interacting hydrologic and geomorphic processes, e.g. precipitation, evapotranspiration, infiltration, RUNOFF GENERATION, EROSION, transport and sedimentation. Watershed management is the administration and regulation of the aggregate resources of a drainage basin to control and regulate water quantity and quality driven by these processes, e.g. for the production of water and the control of erosion, stream flow and floods. Watershed managers use Geographical Information Systems (GIS) and a DIGITAL ELEVATION MODEL (DEM) to delineate watershed characteristics such as watershed boundary, drainage network and contributing subwatersheds for inventory and planning purposes. They also process geo-spatial information on weather and climate, topography, soils and geology, vegetation and land use as well as infrastructural data processed in a GIS to link them with environmental models to analyse and simulate the complex dynamic hydrologic and geomorphic processes for decision-making purposes (see MODELS).

References

Bates, R.L. and Jackson, J.A. (ed.) (1980) *Glossary of Geology*, Falls Church, VA: American Geological Institute.

Fairbridge, R.W. (ed.) (1968) *Encyclopedia of Geomorphology*, New York: Reinhold.

Further reading

Brooks, K.N., Ffolliott, P.F., Gregersen, H.M. and De Bano, L.F. (1997) *Hydrology and the Management of Watersheds*, Ames: Iowa State University Press.

Goodchild, M.F., Steyaert, L.T., Parks, B.O., Johnston, C., Maidment, D., Crane, M. and Glendinning, S. (ed.) (1996) *GIS and Environmental Modeling: Progress and Research Issues*, New York: Wiley.

White, I.D., Mottershead, D.N. and Harrison, S.J. (1984) *Environmental Systems: An Introductory Text*, London: Unwin Hyman.

SEE ALSO: channel, alluvial; drainage basin; hillslope, form; hillslope, process

CHRIS S. RENSCHLER

WAVE

Water waves are periodic undulations of the air-water interface (Figure 174) defined by their height (H), wavelength (L), period of oscillation (T) and speed of propagation (C). They can be *progressive* (e.g. *wind waves, tsunami*) or *standing* (e.g. lake *seiche*, ocean *tide*) and either actively *forced* (e.g. *sea*, tide) or propagate freely (e.g. *swell, tsunami*). The tide is forced by the periodic tractive forces generated by the gravitational attraction of the moon–sun system on the Earth's hydrosphere. TSUNAMI (Japanese for harbour waves) or seismic sea waves result from disturbance of the ocean by an earthquake, an explosive volcanic eruption or a mass failure of ocean sediments. Wind waves are found wherever wind blows over a significant body of water (e.g. marine or lacustrine) and constitute the most important global source of water wave energy. Wind wave size (e.g. height, length, period) depends upon the wind speed, the length of open water over which the wind blows (fetch) and the duration of time that the wind blows. When the oscillatory fluid motions under waves interact with a solid boundary an oscillatory boundary layer is formed; thus, all waves have the potential to generate stresses at the boundary and carry out geomorphological work.

Wave spectrum

By convention, complex irregular wave fields are analysed using Fourier Analysis, which provides a 2-dimensional spectrum of heights and periods. If the direction of wave propagation is included a 3-dimensional *directional spectrum* is produced (Figure 175). Water waves are classified by their frequency (1/T) or period of oscillation, the generating force and the force restoring equilibrium of the water surface (see Kinsman 1965).

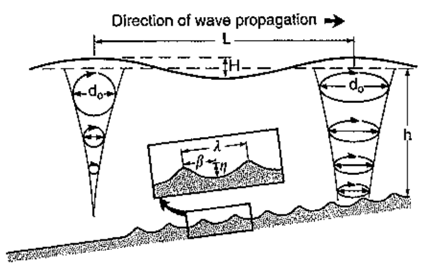


Figure 174 Schematic of wave form and water particle motion in deep and shallow water, where the wave form interacts with a rippled bed. Note: L = wave length; H = wave height; d_o = orbital diameter; h = water depth

TIDES

Tides are the characteristic semi-diurnal or diurnal rise (*flood*) and fall (*ebb*) of the Earth's mean sea level, caused by the periodic components of the gravitational tractive forces generated by the moon and sun; there are at least 360 tidal components leading to modulations of the water surface. Tidal waves are forced standing waves in the world's oceans that are resolved by Coriolis into a series of *rotary standing waves* (Kelvin waves) around the continental margins. These *amphidromic systems* have a central *amphidromic point* of no sea-level change and a *tidal range*, which increases outward to a maximum at the shoreline. Lines of equal tidal range are denoted by co-range lines, which encircle the amphidromic point. Tidal propagation is denoted by *co-tidal lines* (denoting points of equal *phase lag*), which radiate out from the amphidromic point. Such systems occur on all ocean coasts, even where tides are small. Tides are modulated at a range of timescales depending upon the changing relative positions of the Earth, moon and sun. Tides are largest (*spring tides*) when the sun, the moon and the Earth are in alignment (*conjunction* produces slightly larger tides than *opposition*); tides are smallest (*neap tides*) when the moon is in quadrature. Tidal waves are strongly influenced by interaction with the coastal boundary in terms of both their *amplitude* and *phase*, as tides attempt to penetrate into the embayments and estuaries around the coast where friction is important. In the Bay of Fundy, Canada, the tidal range may reach 18 m at springs, since the semi-diurnal forcing by the primary lunar component (M_2) is close to the *natural period of oscillation* of the basin and the shape of the basin causes a distinct topographic forcing. Tides are classified by range: (1) macrotidal >4 m; (2) meso-tidal 2–4 m; (3) micro-tidal <2 m (Davies 1973; Davis and Hayes 1984). *Surges* or meteorological tides are super-elevated tide levels (up to several metres) resulting from the combined effect on the water surface of barometric pressure differentials, wind stress on the water surface and wind wave set-up.

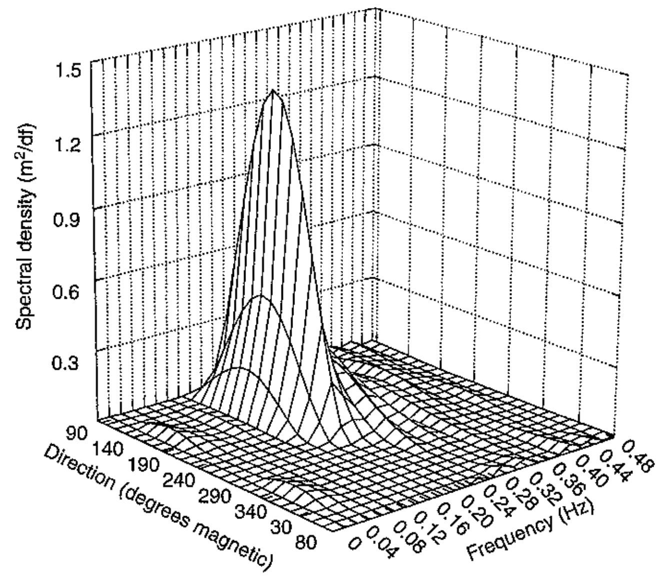


Figure 175 Directional wave spectrum recorded from southern Lake Huron, Canada in ~8 m water depth ~2 km offshore, recorded using a Falmouth Scientific 3DACM Wave Recorder. The record represents a 20-min sample of water surface elevation collected at 2 Hz. Note: a standard frequency spectrum can be obtained by simply projecting the data onto the left-hand frame. These waves have a peak period of ~5 s and a peak direction of approach between 160° and 170°

The rise and fall of mean sea level is complemented by horizontal translations of water known as *tidal currents* or *tidal streams*. Because of the long wave period, the currents are generally small, $<0.05\,m\,s^{-1}$; however, where topographic forcing occurs (e.g. estuaries, straits, inlets, etc.), tidally currents can reach speeds up to $6\,m\,s^{-1}$ (see CURRENTS).

TSUNAMI

In the deep ocean tsunami may be only a few centimetres or decimetres in height and may be difficult to observe, but may have periods ranging from 10–200 min, and hence extremely long wave lengths. Hence they can reach speeds of 700–900 km h^{-1}. In shallow water, as a result of wave *shoaling*, tsunami can reach catastrophic proportions, with run up heights of 10–20 m at the shoreline. The large run-up and large current speeds produce extensive flooding, damage to shorelines and man-made structures and even loss of human life.

GRAVITY AND INFRAGRAVITY WAVES

The most important waves outside the zone of shallow-water wave breaking (*surf zone*) are gravity waves (forced by wind and restored by gravity), with periods ranging from 1–25 s. A modulation of wave height is common, and gravity waves propagate as groups of large waves separated by several smaller waves (Figure 176). This modulation forces a secondary wave, the group-bound long wave, which propagates at the group speed and has a period equal to the group period. Waves with periods of ~25 s to several minutes or longer (frequencies of 0.004–0.04 Hz) are infragravity waves (first recognized and called surf beat by Munk 1949). They are most important to water circulation and sediment transport within the surf zone (Bowen and Huntley 1984). In intermediate water depths, infragravity energy results from wave–wave interactions and consists of a mixture of forced and free wave motions (Herbers *et al.* 1995). Several mechanisms associated with wave breaking in the surf zone have been proposed for their

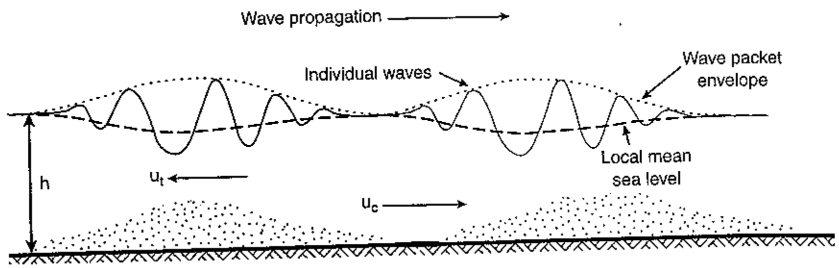


Figure 176 Wave groups and the forced group-bound long wave

generation including: (a) release of the group bound long wave (List 1992); (b) periodic shift in the position of wave breaking (Symonds *et al.* 1982); (c) persistence of groupiness into the surf zone causing a rise and fall of water at the shoreline (Watson and Peregrine 1992). Nearshore infragravity energy can take several forms including standing and progressive *edge waves* (trapped to the shoreline by refraction) and long *leaky waves* (reflected seaward without trapping; Huntley 1976). An important characteristic of infragravity waves in the surf zone is that they increase in amplitude and become more energetic as the offshore wave height increases. In contrast, gravity waves are saturated and thus decrease in importance proportionally during storms. *Shear waves* are a subset of the infragravity field (periods range from several tens of seconds to several tens of minutes) and they form within the surf zone through instabilities in the longshore current. They appear as large fluctuations in the mean current (Bowen and Holman 1989; Oltman-Shay *et al.* 1989) and can contribute significantly to both the cross-shore and alongshore transport of sediment (e.g. Aagaard and Greenwood 1995).

Gravity wave propagation and energy dissipation

Swift (1976) defined a number of *hydraulic provinces* stretching from the deep ocean to the coastline; in the innermost coastal zone (*shoreface*), both the hydrodynamics, sediment transport and morphodynamics are controlled primarily by gravity waves. Propagating gravity waves transfer energy and momentum from offshore to the shoreline. As they move into shallow water, frictional resistance results in a deformation of the orbital motions (Figure 174), a decrease in wave speed and wavelength, and an increase in wave height; this is called *shoaling*. The *specific energy density* (E, according to linear theory = $1/8\ \rho_f\ g\ H^2$, where ρ_f is fluid density, g is the gravitational constant) increases up to a point at which the wave becomes unstable and breaks. During propagation, wave energy may be redistributed laterally through convergence or divergence of the wave *orthogonals*; this is *refraction*. Further, the near-sinusoidal wave form characteristic of deep water becomes increasingly asymmetrical about the horizontal and vertical axes (*wave skewness* and *wave asymmetry*; Figure 177). These non-linearities are critical to the net transport of water and sediment by waves, as they introduce non-symmetrical motion in the oscillatory velocities. In very shallow water, waves break when $H/h \approx 0.4–1.0$. The wave may be destroyed in a *plunging* breaker, producing a large *roller vortex* of the same order as the water depth, or the wave may continue to propagate as a *spilling* breaker, with collapse of the leading face of the wave. Surf *bores* are solitary waves, propagating across the surf zone controlled only by the water depth. The surf zone is a complex hydrodynamic environment, where interactions occur between incident waves, macro and micro turbulent vortices, secondary waves and quasi-steady currents of various origins (for a review see Kobayashi 1988). Ultimately wave energy is either reflected from the beach or dissipated in the surf zone and the reversing *swash* currents on the beach face. Sediment transport in the *uprush* and *backwash* is constrained by extremely small water depths, large Froude Numbers, large pressure gradients near *bore* fronts and the *infiltration* or *exfiltration* of water from the beach face *water table* (for recent review see Butt and Russell 2000).

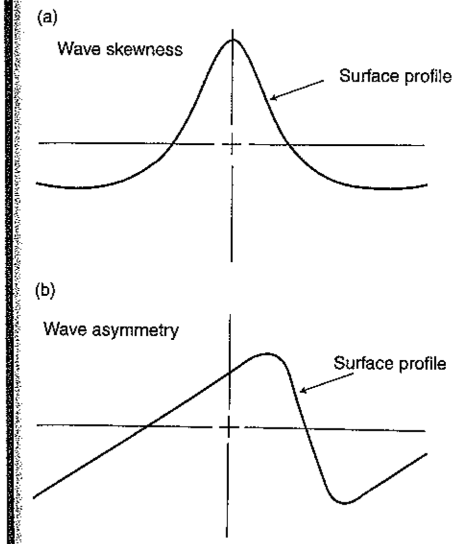


Figure 177 Asymmetrical wave forms: (a) asymmetry about the horizontal axis – wave skewness; (b) asymmetry about the vertical axis – wave asymmetry

Stresses generated by waves

Stresses generated by wave motion can be considered to take two forms: (a) *radiation stress* is the excess flux of momentum due to the presence of waves in the water and operates between the wave crest and trough. The concept was introduced by Longuet-Higgins and Stewart (1964). It explains the generation of the group-forced long wave in deep water and a number of hydrodynamic phenomena in the surf zone, such as wave *set-up* and *set-down* (i.e. the raising and lowering of the still water level at the shoreline), the *near-bed return flow* known as *undertow; longshore currents* (see CURRENTS); (b) the *boundary layer stresses* under waves are significantly larger than those of an equivalent quasi-steady current. Since the boundary layer develops and decays each half-wave cycle, it is much thinner and the resultant velocity gradients much larger. The time-varying bed shear stresses (τ_w), which have units of force per unit area (N m^{-2} in SI units) can also be written in units of velocity using the friction or shear velocity (u_{*w} which is not a real velocity, but is used for convenience and can be related to the fluid turbulence) and the fluid mass density (ρ_f):

$$\tau_w = \rho_f {u_*}_w^2$$
$$u_{*w} = (\tau_w/\rho_f)^{1/2}$$

A dimensionless form of the bed shear stress is given by the *Shields Parameter* (θ_w):

$$\theta_w = \tau_w/\{g\ (\rho_s - \rho_f)\ D\}$$

where D is the particle size on the bed. The bed shear therefore depends not only on the flow velocity but also the *bed roughness* (grain size and bedforms). When quasi-steady currents are combined with waves (usually the case), these relationships become more complex (Soulsby 1997). However, if during the wave cycle these stresses exceed that for the *initiation of motion* of particles on the bed, entrained material will begin to move and oscillate in place. However, it is rare that a perfect oscillatory motion occurs in nature and rare to find a perfectly horizontal bed where gravity has no effect; thus a net displacement of particles occurs. Usually, there is also superimposed quasi-steady current, either induced by *wave streaming* (mass transport), the *asymmetry* of the waveform, or by some other process, such as *undertow* or *longshore currents*, which will enhance the net motion.

Sediment transport by waves

A simple concept of waves 'stirring' and currents 'transporting' sediment has been used to model transport under waves (Bagnold 1966; Bowen 1980; Bailard 1981); however this is not strictly correct. Waves and currents interact in a complex, non-linear manner and there is strong feedback from the bed materials (particle size, bedforms, bed slope, etc.). This is especially true for sediment transport, where rates have been related to some power of the horizontal velocity, and exponents varying from 3 to 7 have been used (Soulsby 1997). In the presence of ripples or megaripples, there is a complex, periodic shedding of turbulent sediment-laden vortices formed by flow separation at the sharp discontinuity of the bedform crest (Sleath 1984). In general these vortices are released upwards into the flow close to the time of flow reversal, when the oscillatory currents drop to zero. However, the timing of this release is critical as the vortices can be advected either down wave or upwave depending upon the release time and the wave period.

Geomorphological response to waves

The geomorphological response to waves depends on a range of variables including: (a) the timescale; (b) the spatial scale; (c) the geological and bathymetric setting; (d) the tidal range; etc. Forms can range from small sand ripples of a few centimetres in height and perhaps 10 cm in spacing, to the upwardly concave shape of the shoreface profile out to the depth at which waves begin to shoal (called *wave base*). Coastal morphologies are often classified by the nature of the substrate: (a) *rock, cliff* or *cohesive* coasts (see Tsunamura 1992; Trenhaile 1987); (b) *sedimentary* coasts (see Davis 1985), where the grain size is a critical determinant of form (gravel, sand or mud coasts). Sedimentary coasts account for approximately 20 per cent of the world's coastline, of which sandy coasts make up 10–15 per cent, and rocky coasts make up approximately 80 per cent. The vast majority of wave studies have been restricted to sedimentary, especially sandy, coasts, where measurable change is relatively rapid compared to rocky coasts.

ROCK, CLIFF AND COHESIVE COASTS

On bedrock and cohesive coasts exposed to large *wind waves* erosion of the base of cliffs (see CLIFF, COASTAL) or bluffs may occur, producing an indentation or *wave-cut notch*. The material above the notch may become unstable and *mass wasting* processes will induce failure. The resulting debris is then entrained by the waves and transported alongshore and offshore (depending on grain size) by wave-generated currents. The factors governing basal erosion rates are: (1) the *hydraulic* and *pneumatic* action of the waves, modulated by the water level, the upper shoreface slope and any beach materials present; (2) the *resistance* of the bedrock, which is controlled by its lithology, structure and geotechnical properties (Tsunamura 1992). The hydraulic action of waves consist of: (a) *compressive* forces, which vary with the waveform (standing wave, breaking wave, broken wave) at the cliff toe; the maximum compressive pressure is exerted by a wave that breaks exactly at the cliff face; (b) *pneumatic* forces may be induced by compression of air by waves in fractures and joints or other indentations and its explosive expansive release as the wave recedes; (c) *shearing* and *tension* forces develop across the rock surface as the wave uprush and backwash generate boundary forces capable of entraining loosened debris. The total mechanical process is known as *quarrying* or *plucking*. When significant sediment load is available, waves also cause erosion through *abrasion* and *impact* forces by the mobilized sedimentary particles. No field measurements have yet been made of the forces acting against a cliff and most of the information comes from theoretical calculations or laboratory simulations. Major questions concerning rock coasts revolve around the relative importance of waves and tides and subaerial weathering, as well as the extent to which forms may have been inherited from a previous time (Trenhaile 2002). Bryant and Young (1996) cite evidence for erosion of rock cliffs up to 15 m above present sea level by *tsunami* (see also Aalto *et al.* 1999).

Typical coastal forms are: (a) *steep cliffs*; (b) *wave-cut notches*; (c) *sea caves*; (d) *rock arches* and *stacks*; and (c) *shore platforms*, with or without a seaward *rampart*. All involve wave erosion to a greater or lesser degree. Mass wasting processes, such as *rock falls* and *planar* or *rotational slides*, *earth* and *debris flows*, and *piping* play a major role in maintaining cliffs along both rocky and cohesive coasts. Weathering processes are also critical in making material available for wave erosion and transport. The most characteristic form is a steep cliff and shore platform, with the latter being gently sloping or subhorizontal with a sharp break at the seaward end. Here the tide plays a major role in translating the hydrodynamic zones horizontally, so controlling the location, duration and type of wave action which occurs, but it plays no active role.

SEDIMENTARY COASTS

In cohesionless and very weakly cohesive materials, the shoreface is generally upwardly concave with slopes increasing toward the shoreline. Associated with this is a general shoreward increase in grain size. The increasing slope generates gravitational forces on the sediment to balance the increasing stresses from the landward propagating waves. An equilibrium profile shape was defined empirically by Dean (1991):

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

where h = water depth, A = a constant and x is the distance offshore. Using the suspended sediment transport model of Bagnold, Bowen (1980) showed that if the equilibrium profile is defined as that shape where the net sediment transport over a wave period is everywhere zero, then the exponent 2/3 also applied. Van Rijn (1998) gives an extensive review of the literature on the *profile of equilibrium* concept.

Once waves break in the nearshore, the force balance on the bed is no longer simple. The occurrence of increased turbulence (macro and micro), secondary waves at a range of frequencies, secondary quasi-steady currents, and alongshore and cross-shore pressure gradient currents lead to complex sediment flux patterns, morphological forms and highly variable grain size. On many sandy coasts the upper shoreface profile is characterized by from 1 to 13 sand ridges called bars, which may be 2 or 3-dimensional, shore-parallel or shore-normal and, in the extreme case, form classic crescentic patterns (Greenwood 2003). The generation of barred profiles, their spatial and temporal dynamics have been the focus of extensive research. In 1979, Greenwood and Davidson-Arnott classified bars by their location and the primary forcing agents. Wright and Short (1984) proposed that nearshore bars (their number, form, etc.) are simply sequential changes as wave energy increases and decreases over time from a fully dissipative beach state to a fully reflective state. Using Dean's dimensionless fall velocity parameter (Ω), based on wave height (H_b) at breaking, wave period (T) and sediment size (settling velocity, w_s):

$$\Omega = H_b/(w_s T)$$

Wright and Short (1984) linked this to six distinct beach states; a similar *model* sequence was proposed by Lippmann and Holman (1990). Recently, attention has been directed at the feedback between the shoreface topography and the hydrodynamics, leading to the concept of self-organization of morphological forms (Plant *et al.* 2001; Wijnberg and Kroon 2002).

References

Aagaard, T. and Greenwood, B. (1995) Suspended sediment transport and morphological response on a dissipative beach, *Continental Shelf Research* 15, 1,061–1,086.

Aalto, K.R., Alto, R., Garrison-Laney, C.E. and Abramson, H.F. (1999) Tsunami(?) sculpturing of the Pebble Beach wave-cut platform, Crescent City area, California, *Journal of Geology* 107, 607–622.

Bagnold, R.A. (1966) An approach to the sediment transport problem from general physics, *US Geological Survey, Professional Paper* 422-I.

Bailard, J.A. (1981) An energetics total load sediment transport model for a plane sloping beach, *Journal of Geophysical Research* 86, 10,938–10,954.

Bowen, A.J. (1980) Simple models of nearshore sedimentation: beach profiles and longshore bars, in S.B. McCann (ed.) *The Coastline of Canada*, 1–11, Geological Survey of Canada, Paper 80–10.

Bowen, A.J. and Holman, R.A. (1989) Shear instabilities of the mean longshore current, 1. Theory, *Journal of Geophysical Research* 94, 18,023–18,030.

Bowen, A.J. and Huntley, D.A. (1984) Waves, long waves and nearshore morphology, *Marine Geology* 60, 1–13.

Bryant, E.A. and Young, R.W. (1996) Bedrock-sculpturing by tsunami, south coast, New South Wales, Australia, *Journal of Geology* 104, 565–582.

Butt, T. and Russell, P. (2000) Hydrodynamics and cross-shore sediment transport in the swash zone of natural beaches: a review, *Journal of Coastal Research* 16, 255–268.

Davies, J.L. (1973) *Geographical Variation in Coastal Development*, New York: Hafner.

Davis, R.A. Jr (ed.) (1985) *Coastal Sedimentary Environments*, New York: Springer-Verlag.

Davis, R.A. Jr and Hayes, M.O. (1984) What is a wave-dominated coast, *Marine Geology* 60, 313–329.

Dean, R.G. (1991) Equilibrium beach profiles, *Journal of Coastal Research* 7, 53–84.

Greenwood, B. (2003) Wave-formed bars, in M. Schwartz (ed.) *Encyclopedia of Coastal Science*, Amsterdam: Kluwer (in press).

Greenwood, B. and Davidson-Arnott, R.G.D. (1979) Sedimentation and equilibrium in wave-formed bars: a review and case study, *Canadian Journal of Earth Sciences* 16, 312–332.

Herbers, T.H.C., Elgar, S. and Guza, R.T. (1995) Generation and propagation of infragravity waves, *Journal of Geophysical Research* 100, 24,863–24,872.

Huntley, D.H. (1976) Long period waves on a natural beach, *Journal of Geophysical Research* 81, 6,441–6,449.

Kinsman, B. (1965) *Wind Waves: Their Generation and Propagation on the Ocean Surface*, Englewood Cliffs, NJ: Prentice Hall.

Kobayashi, N. (1988) Review of wave transformation and cross-shore sediment transport processes in surf zones, *Journal of Coastal Research* 4, 435–445.

Lippmann, T.L. and Holman, R.A. (1990) The spatial and temporal variability of sand bar morphology, *Journal of Geophysical Research* 95, 11,575–11,590.

List, J.H. (1992) A model for the generation of two-dimensional surf beat, *Journal of Geophysical Research* 97, 5,263–5,635.

Longuet-Higgins, M.S. and Stewart, R.W. (1964) Radiation stress in water waves, a physical discussion with applications, *Deep Sea Research* 11, 529–563.

Munk, W.H. (1949) Surf beats, *Transactions American Geophysical Union* 30, 849–854.

Oltman-Shay, J., Howd, P.A. and Birkemeir, W.A. (1989) Shear instabilities of the mean longshore current, 2. Field observations, *Journal of Geophysical Research* 94, 18,031–18,042.

Plant, N.G., Freilich, M.H. and Holman, R.A. (2001) The role of morphologic feedback in surf zone sand bar response, *Journal of Geophysical Research* 106, 959–971.

Sleath, J.F. (1984) *Seabed Mechanics*, New York: Wiley.

Soulsby, R. (1997) *Dynamics of Marine Sands*, London: Thomas Telford.
Swift, D.J.P. (1976) Coastal Sedimentation; in D.J. Stanley and D.J.P. Swift (eds) *Marine Sediment Transport and Environmental Management*, 311–350, New York: Wiley Interscience.
Symonds, G., Huntley, D.A. and Bowen, A.J. (1982) Two-dimensional surf beat: long wave generation by a time-varying breakpoint, *Journal of Geophysical Research* 87, 492–498.
Trenhaile, A.S. (1987) *The Geomorphology of Rock Coasts*, Oxford: Oxford University Press.
——(2002) Rock coasts, with particular emphasis on shore platforms, *Geomorphology* 48, 7–22.
Tsunamura, T. (1992) *Geomorphology of Rocky Coasts*, Chichester: Wiley.
Van Rijn, L.C. (1998) *Principles of Coastal Morphology*, Amsterdam: Aqua Publications.
Watson, G. and Peregrine, D.H. (1992) Low frequency waves in the surf zone, *Proceedings of the Twenty-third International Conference on Coastal Engineering*, American Society of Civil Engineers, 818–831.
Wijnberg, K.M. and Kroon, A. (2002) Barred beaches, *Geomorphology* 48, 103–120.
Wright, L.D. and Short, A.D. (1984) Morphodynamic variability of surf zones and beaches: a synthesis, *Marine* Geology 56, 93–118.

Further reading

Komar, P.D. (1998) *Beach Processes and Sedimentation*, 2nd edition, Upper Saddle River, NJ: Prentice Hall.
Open University (1989) *Waves, Tides and Shallow-water Processes*, Oxford: Pergamon Press.
Trenhaile, A.S. (1997) *Coastal Dynamics and Landforms*, Oxford: Oxford University Press.

BRIAN GREENWOOD

WEATHERING

Weathering refers to a group of processes collectively responsible for the breakdown of materials at or near the Earth's surface. Weathering occurs because the environmental conditions under which most rock materials formed differ substantially from those which prevail near the Earth's surface. Consequently, they undergo a variety of modifications which result in more stable products under the newly imposed conditions of temperature, pressure, moisture and gaseous environment.

From a geomorphological perspective, rock weathering is extremely important. First, weathering processes prepare Earth materials for subsequent transportation by agents of erosion. Second, weathering is an essential component of soil formation at the Earth's surface. Third, weathering processes are directly responsible for land form and landscape evolution. Karst landscapes and their distinctive land form assemblages, for example, are a direct consequence of weathering processes, as are the thick regolith-mantled landscapes of the tropics and subtropics.

Weathering processes

Traditionally, weathering processes are regarded as being physical, chemical or biological in nature. In reality, the three groups of processes act synergistically making it difficult to distinguish clearly between the effects of any single one (Pope *et al.* 1995). For instance, biological processes effect both physical and chemical change within the weathering system.

PHYSICAL WEATHERING

Physical weathering collectively involves the breakdown of rock material into smaller pieces without any change in the chemistry or mineralogy of the rock. Rock material is simply disaggregated as a result of the effects of the generation of forces within the rock mass. These forces are derived either from internal expansion of rocks and minerals, or from growth of materials in voids. The principal physical weathering processes associated with internal expansion include: INSOLATION WEATHERING (thermal expansion), unloading, HYDRATION, WETTING AND DRYING WEATHERING, organic expansion and SALT WEATHERING. Growth in voids is dominated by frost action and growth of salt crystals.

Insolation weathering refers to the breakdown of rock as a result of expansion and contraction caused by frequent temperature changes. As bedrock has low thermal conductivity, the surface of rock material expands more than the interior of the rock and, as a result, stresses are set up that eventually lead to the disintegration of the rock. Its effectiveness is enhanced when rocks consist of a mixture of light and dark minerals thus facilitating marked variation in thermal conductivity with individual minerals expanding and contracting at different rates. Research from engineering and ceramics has demonstrated that relatively small temperature variations over short time periods can most effectively disintegrate rock. Thermal stress has also been shown to be intimately associated with the breakdown of rock in areas affected by fire (Ollier and Ash 1983).

Unloading, or sheeting, refers to the formation of slabs of rock parallel to the ground surface, but separated from the underlying intact rock by joints. Unloading or sheet joints are generally attributed to the reduction in compressive stress in rock masses as a result of erosion of the upper part of the rock mass which then fails in the direction of stress removal. Failure is likely to begin with the extension of an initial crack and continue with the development of sheet joints parallel to the unloading surface. While this process appears to be most effective in brittle crystalline rocks it is also widely developed in massive sandstones.

While hydration begins as fundamentally a chemical process, with the absorption of water along fractures, especially cleavage planes in minerals, its effect is primarily a physical one. Expansion of minerals as a result of the incorporation of water into the crystal lattice can result in the production of tremendous force, especially in confined spaces. The effect of mineral expansion is to disrupt adjacent mineral grains causing loss of grain boundary cohesion and ultimately bedrock disaggregation to produce friable rubble referred to as grus. Similar processes concentrated around the edges of joint blocks where water is most abundant leads to the formation of concentric shells of weathered material referred to as spheroidal weathering or EXFOLIATION. The formation of these concentric shells of weathered material is also sometimes referred to as onion skin weathering or desquamation.

Repeated wetting and drying of rock may be a significant physical weathering process. Rock breakdown by wetting and drying involves the development of internal rock stresses as a result of the progressive formation of ordered water layers which exert forces against confining walls or void boundaries. As the positively charged ends of water molecules are attracted to the negatively charged surfaces of clay minerals and colloids they form a layer of oriented water particles. With each wetting event a new layer of ordered water is added to clay particles, which remain during the drying phase. Failure appears to be most pronounced during the drying phase when negative pore-water pressure is greatest and tensile failure occurs.

Frost weathering (see FROST AND FROST WEATHERING) has long been viewed as one of the most significant physical weathering processes in cold climates. Traditionally, frost weathering has been believed to occur as a result of forces generated in association with the volume increase accompanying the phase change from liquid water to ice. However, the theoretical assumptions underlying the generation of such forces are seldom met in nature. A more realistic model, with growing empirical support, is the segregated ice model of frost weathering in which expansion is the result of the migration of unfrozen water toward growing ice lenses and only secondarily the result of volumetric change accompanying phase change. Frost weathering is thereby the result of enlargement of microcracks and relatively large pores by ice growth accompanying ice segregation (Walder and Hallet 1986).

The growth of salt crystals in voids in rock exerts forces capable of disaggregating rock material when the tensile strength of the rock is exceeded. Salt crystal growth occurs when solutions containing salts are evaporated or in some cases cooled, when water is added to salts and hydration occurs and when salts are heated. All these processes result in substantial increases in volume of salt crystals and accompanying application of force against the walls of voids, leading ultimately to rock disruption and disintegration (Goudie 1997).

Biological organisms are also effective agents of mechanical weathering. Biophysical weathering processes include root wedging, and lichen, algal and bacterial activity. It has been widely suggested that the penetration of plant roots along fractures in rocks is capable of splitting rocks apart as the root grows. Physical weathering by lichens is accomplished in two principal ways. First, by the penetration of hyphae along microcracks and grain boundaries. The penetration and growth of hyphae have been shown to exert tensile stresses which exceed the tensile strength of most rocks. Second, by the expansion of lichen thalli and hyphae due to the uptake of moisture. Such moisture uptake substantially increases the size of the thallus and hyphae thus exerting considerable pressure on the rock surface. The efficiency of this process is seen in the frequent incorporation of rock fragments into the thallus (Barker *et al.* 1997). Algae also appear to be important contributors to the breakup of rock materials. It has been found that upon wetting, the polymer sheath surrounding algal cells increases by as much as twenty times, thus exerting expansion forces sufficient to pry already weakened rock flakes from a rock surface (Hall and Otte 1990).

CHEMICAL WEATHERING

CHEMICAL WEATHERING processes are those which involve the chemical and or mineralogical transformation of rocks and minerals at/or near the Earth's surface into products that are in equilibrium with Earth surface conditions. Several principal chemical weathering processes are recognized, including solution (dissolution), HYDROLYSIS, HYDRATION, carbonation, chelation and redox reactions. Solution refers to the dissolving of minerals in the presence of water and involves the removal of atoms from mineral surfaces thus reducing the stability of minerals and enhancing their vulnerability to subsequent chemical degradation (Blum and Stillings 1995). Hydrolysis refers to the reaction of hydrogen in solution with mineral surfaces and subsequent formation of secondary clay minerals as displaced cations react with hydroxy ions in adhered water on mineral surfaces. Hydration involves the addition of water to a mineral structure to form a new mineral. Redox reactions are reduction/oxidation reactions which involve reactions with oxygen in the atmosphere. OXIDATION involves a loss of electrons while REDUCTION involves a gain of electrons: oxidation of one mineral component is achieved through reduction of another. Iron is the most commonly affected chemical species. Carbonation involves the reaction of minerals with carbon dioxide in the presence of water and is especially important in the chemical weathering of limestones and other carbonate-rich rocks. In the soil environment, where the weathering system is dominated by clay mineral–soil solution interactions, ion exchange reactions occur. Ion exchange involves the transfer of ions between solution and mineral, and generally involves the replacement of ions in the mineral interlayer though replacement within the crystal lattice can also occur. The efficiency of this process is to a large degree controlled by the strength of electrical double layer, an exposed plane of negatively charged oxygen ions and the balancing swarm of exchangeable cations (Jenny 1980).

Plants represent significant contributors to the chemical weathering of rocks and minerals, primarily due to the fact that they produce abundant quantities of organic acids. These organic acids form ring structures around a metal core with multiple bonds between the organic acid and the metal. They are responsible for the chelation of cations such as Fe and Al. In addition, during their formation they commonly release H^+ which further reacts with mineral surfaces. Most of these acids are produced in the vicinity of root tips.

Lichens, algae and bacteria produce abundant organic acids which are responsible for considerable rock and mineral weathering. These organisms are responsible for the production of two principal groups of organic compounds. These include oxalic acid and various oxalates. The former compound possesses high solubility and contributes abundant protons for mineral dissolution as well as producing ring structures for chelation. Oxalates have been shown to enhance mineral grain dissolution through proton donation (Barker *et al.* 1997).

Weathering intensity

Weathering intensity refers to the degree of decomposition of a rock or mineral; it is a measure of the amount of alteration that has occurred. Numerous indices of weathering intensity have been developed and include both descriptive measures which are based on changes in the visual appearance of regolith as it becomes more altered, as well as quantitative measures of the amount of chemical and/or mineralogical change that has occurred to the primary mineral or unaltered rock. Descriptive or qualitative measures of weathering intensity have been widely used in the fields of engineering geology and regolith geomorphology. They typically are descriptive and subjective, based on changes in visual appearance of materials associated with progressive disaggregation or loss of cohesion. These classifications typically contain a limited number of weathering intensity categories, generally consisting of categories such as fresh rock at the unweathered end and soil at the most weathered end with slightly, moderately, highly and completely weathered categories in between. The boundary between the fresh, unaltered bedrock and altered material is referred to as the WEATHERING FRONT. This boundary may be quite abrupt, but more commonly it is highly irregular.

Semi-quantitative measures of weathering intensity involve the assessment of the relative hardness of fresh and various states of altered material. The most widely used of these methods is the SCHMIDT HAMMER which provides an index of hardness based on the amount of resistance to the compressive stress applied to the rock by the hammer.

A large number of quantitative measures of the intensity of weathering are available that are based on chemical and mineralogical characteristics of the weathering rocks. In virtually all cases, these indices compare abundances of non-resistant constituents to resistant constituents. They are expressed as dimensionless ratios which generally decrease as the intensity of weathering increases: as more non-resistant materials are removed. The more commonly used ratios include silica:iron, silica:aluminium, silica:sesquioxides, silica:resistates, bases:alumina, alkalis:resitates, and alkali earths: resistates. In addition several more comprehensive chemical weathering indicies have been developed and still receive limited use including the Parker Weathering Index, and the Reiche Weathering Potential Index.

Several mineral weathering indices exist such as quartz:feldspar ratios and multiple-mineral weathering indices including those for heavy minerals such as zircon + tourmaline:amphiboles + pyroxenes which increase as weathering increases. Several methods involving the characteristics of individual mineral grains have also been developed for evaluating the intensity of weathering including surface micro-textural features of heavy minerals and the degree of etching on ferromagnesian minerals.

The intensity of weathering is influenced by numerous interacting factors which affect both the extent to which primary minerals and fresh rock have been altered as well as the rate at which alteration takes place. Intrinsic factors include the mineralogy and chemistry of the parent material, grain size and shape, porosity, and fracture abundance and openness. Extrinsic factors generally include temperature, moisture abundance and water chemistry. Traditionally it is the role of temperature and moisture that have been emphasized in controlling both rates and intensity of weathering at the global scale as portrayed in the weathering models of Peltier (1950) who used temperature and precipitation as a basis for a global model of the variability of physical and chemical weathering processes. Similarly, Strakhov (1967) portrays the arid/semi-arid and tundra regions of the world as possessing a weathering mantle dominated by no more than disaggregated bedrock with essentially no chemical or mineralogical alteration, while the hot, wet tropics/subtropics carry weathering mantles dominated by the concentration of sesquioxides. In fact, the intensity of weathering is not controlled by tropospheric climate, rather it is controlled by a combination of boundary layer and reaction site temperature and moisture. More specifically, the intensity of chemical weathering is controlled by a complex set of synergistically operating factors that control both weathering intensity and rate including: the availability and proximity of both biotic and abiotic weathering agents; mineralogy, chemistry and petrology of the parent material; structure of the parent material at multiple scales; temperature at the reaction site; hydraulics of water movement; removal of weathered materials; addition/removal of organic and inorganic fines; microtopography of both landscape surface and mineral surface; exposed surface area and the presence of any accreted surface coatings (Pope *et al.* 1995).

Deep weathering

In many parts of the world, weathering profiles reach extraordinary thicknesses. These weathering profiles are commonly referred to as deep weathering profiles. Weathering to depths in excess of 100 m is not uncommon, and cases of weathering profiles extending to a kilometre or more, not unknown. While great depths of weathering have been traditionally recognized from the tropics this does not mean that deep weathering profiles outside the tropics necessarily imply their formation under such climatic conditions. In fact deep weathering profiles can develop in a wide range of climatic settings in which weathering has been prolonged. Deep weathering profiles around the world display a marked variation in the intensity of weathering. On the ancient landscapes of Australia and Africa, deep weathering profiles are characteristically intensely weathered, with the clay mineral kaolinite dominating. These intensely weathered deep profiles simply reflect the availability of abundant moisture and efficient removal of weathered products in solution. It is important to point out that deep weathering profiles displaying little chemical and mineralogical transformation also occur extensively in Australia, Europe and North America. These profiles reflect less efficient leaching and often also reflect weakening of rock strength early in the geologic evolution of the parent material by providing exceedingly deep fracture systems which can be exploited by circulating meteoric waters.

The deep, kaolinite-dominated, weathering profiles of Australia and Africa, commonly are capped by a strongly indurated crust enriched in a variety of chemical cementing agents. Collectively, these indurated crusts are referred to as DURICRUSTS and are most frequently, but not exclusively, cemented by iron, aluminium, silica, calcium carbonate or gypsum.

References

Barker, W.W., Welch, S.A. and Banfield, J.F. (1997) Biogeochemical weathering of silicate minerals, in J.F. Banfield and K.H. Nealson (eds) *Geomicrobiology: Interactions between Microbes and Minerals*, 391–428, Reviews in Mineralogy 35, Washington, DC: Mineralogical Society of America.

Blum, A.E. and Stillings, L.L. (1995) Feldspar dissolution kinetics, in A.F. White and S.L. Brantly, (eds) *Chemical Weathering Rates of Silicate Minerals*, 291–351, Reviews in Mineralogy 31, Washington, DC: Mineralogical Society of America.

Goudie, A.S. (1997) Weathering processes, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 2nd edition, 25–39, Chichester: Wiley.

Hall, K. and Otte, W. (1990) A note on biological weathering of Nunataks of the Juneau Icefield, Alaska, *Permafrost and Periglacial Processes* 1, 189–196.

Jenny, H. (1980) *The Soil Resource: Origin and Behaviour*, New York: Springer Verlag.

Ollier, C.D. and Ash, J.E. (1983) Fire and rock breakdown, *Zeitschrift für Geomorphologie* 27, 363–374.

Peltier, L. (1950) The geographic cycle in periglacial regions as it is related to climatic geomorphology, *Annals of the Association of American Geographers* 40, 214–236.

Pope, G., Dorn, R.I. and Dixon, J.C. (1995) A new conceptual model for understanding geographical variations in weathering, *Annals of the Association of American Geographers* 85, 38–64.

Strakhov, N.M. (1967) *Principles of Lithogenesis*, Edinburgh: Oliver and Boyd.

Walder, J.S. and Hallet, B. (1986) The physical basis of frost weathering: toward a more fundamental and unified perspective, *Arctic and Alpine Research* 18, 27–32.

Further reading

Bland, W. and Rolls, D. (1998) *Weathering: An Introduction to the Scientific Principles*, London: Arnold.

White, A.F. and Brantly, S.L. (eds) (1995) *Chemical Weathering Rates of Silicate Minerals*, Reviews in Mineralogy 31, Washington, DC: Mineralogical Society of America.

Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

JOHN C. DIXON

WEATHERING AND CLIMATE CHANGE

Weathering plays a fundamental role in the carbon cycle and hence influences the role of CO_2 as a greenhouse gas. While volcanoes add CO_2 to the atmosphere, the major long-term process of CO_2 removal is chemical weathering of continental rocks. The prime weathering mechanism involved in the process is HYDROLYSIS. Rates of chemical weathering vary through time in response to changes in temperature and precipitation amounts. They are also affected by vegetation cover, the nature of which is also linked to temperature and vegetation conditions. Thus climate affects the global rate of weathering, but weathering has the capacity of altering the climate by regulating the rate at which CO_2 is removed from the atmosphere. Weathering may act as a negative feedback that moderates long-term climatic change (Figure 178) (Ruddiman 2000).

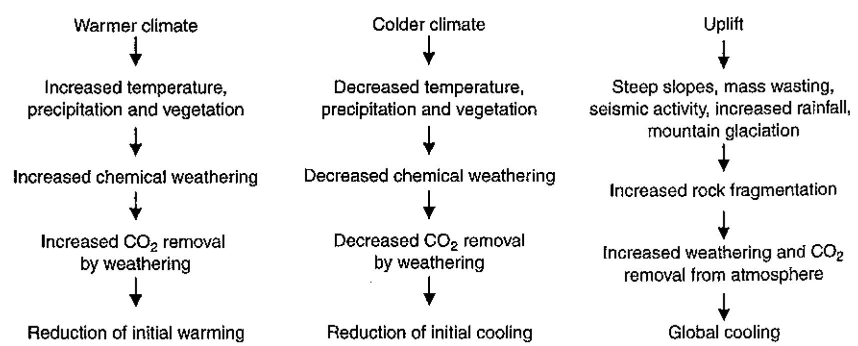


Figure 178 Negative feedbacks between weathering and climate change

However, it is also possible to see chemical weathering not only as a negative feedback that moderates climate changes, but also as an active driver of climatic change. The 'uplift weathering hypothesis' asserts that exposure of fragmented and unweathered rock is an important factor controlling the intensity of chemical weathering. It also asserts that rates of exposure of fresh rock are increased in areas of active mountain building because of seismic activity and mass wasting processes. Furthermore, uplifting areas generate more precipitation and glaciation. Mountain glaciers create pulverized bedrock.

Uplift has been active in the last few tens of millions of years (e.g. the Himalayas, Tibet, Andes, Rocky Mountains, European Alps). It is thus possible that accelerated chemical weathering has contributed to the climatic decline of the Late Cenozoic. It is also possible that differences in rates of chemical weathering between glacial and interglacial conditions in the Pleistocene occurred and contributed to the low CO_2 levels in the atmosphere during cold phases (Munhoven 2002). Karstic dissolution may also be an important component of the global carbon cycle (Gombert 2002).

The relationships between chemical weathering rates and climatic conditions are, however, complex (Kump *et al.* 2000) and the detection of variations in weathering intensity in the geologic record by the study of radiogenic isotope ratios ($^{87/86}$Sr, $^{143/144}$Nd, $^{187/186}$OS) is not without problems.

References

Gombert, P. (2002) Role of karstic dissolution in global carbon cycle, *Global and Planetary Change* 33, 177–184.

Kump, L.R., Bromtley, S.L. and Arthur, M.A. (2000) Chemical weathering, atmospheric CO_2, and climate, *Annual Reviews of Earth and Planetary Science* 28, 611–667.

Munhoven, G. (2002) Glacial–interglacial changes of continental weathering: estimates of the related CO_2 and HCO_3 flux variations and their uncertainties, *Global and Planetary Change* 33, 155–176.

Ruddiman, W.F. (2000) *Earth's Climate, Past and Future*, New York: W.H. Freeman.

A.S. GOUDIE

WEATHERING FRONT

The term pertains to subsurface weathering and is used to describe the boundary that separates solid, unweathered rock and rock that has already been weathered but remains still *in situ*. In reality, the transition between weathered and fresh rock compartments is rarely sharp. Usually there exists a transitional zone, in certain circumstances only a few centimetres thick, within which the change from unweathered mass to disintegrated or decomposed rock takes place. In some lithologies, especially in foliated metamorphic rocks and clastic sedimentary rocks, the transition is highly gradational and a well-defined weathering front may not exist. By contrast, weathering profiles in igneous rocks such as granite tend to have a very clear lower boundary.

The concept of weathering front has evolved from the notion of 'basal platform' introduced by Linton (1955) to explain the origin of TORS. Later it was shown that the boundary is hardly ever planar and the 'basal platform' is better replaced by 'basal surface of weathering' (Ruxton and Berry 1957). This implies a stable position of the rock/weathering mantle boundary and is not adequate to describe weathering profiles with abundant core stones. Therefore the present term 'weathering front' was recommended for use (Mabbutt 1961).

A weathering front is a dynamic feature, which migrates downwards and sidewards over time, as more and more rock disintegrates or decomposes. It should not be visualized as a single, continuous surface present at some depth, as each core stone is separated from the surrounding SAPROLITE by the weathering front. If the thickness of the weathering mantle is highly variable over short distances, the topography of the weathering front will accordingly be very rough.

References

Linton, D.L. (1955) The problem of tors, *Geographical Journal* 121, 470–487.

Mabbutt, J.A. (1961) Basal surface or weathering front, *Proceedings Geologists' Association* 72, 357–358.

Ruxton, B.P. and Berry, L. (1957) Weathering of granite and associated erosional features in Hong Kong, *Geological Society of America Bulletin* 68, 1,263–1,292.

SEE ALSO: deep weathering; etching, etchplain and etchplanation; saprolite

PIOTR MIGOŃ

WEATHERING-LIMITED AND TRANSPORT-LIMITED

Weathering and transport limitations have been used as an important concept in geomorphology and are basic to an understanding of hillslope development. Broadly defined, a weathering-limited system is one in which the supply of material determines the flux of mass, while in a transport-limited system, sufficient material is available at any given time for mass movement to occur.

The two terms were originally introduced to geomorphology by G.K. Gilbert in his monograph on the geology of the Henry Mountains (Gilbert 1877). In this classic work the author refers to two landscape types: one in which the overall rate of DENUDATION remains limited by the rate of rock weathering and mass movement processes are very effective in removing deposited materials as they accumulate (weathering-limited); the second is characterized by weathering processes exceeding the rate at which material can be denuded and there is a net accumulation of rock (transport-limited).

The same concept was revisited by Alfred Jahn in 1954, but remained largely unnoticed until translated from Polish into English (Jahn 1968). Jahn uses the denudational balance of slopes to classify slope processes. Although Jahn does not explicitly use the terms weathering-limited and transport-limited in his 1968 publication, he clearly separates slopes into those where the intensity of weathering is lower than that of material transport and those where the opposite is true. He further explains that the denudational balance is in equilibrium when the intensity of weathering equals that of transport.

Eighteen years after Jahn's original publication, Kirkby (Carson and Kirkby 1972) first coined the terms weathering-limited and transport-limited, leaning on Gilbert's and Jahn's early explanations and defining weathering-limited slopes as those in which the transport processes are more rapid than weathering. In contrast, transport-limited slopes are characterized by weathering rates in excess of transport rates which implies the formation of a soil cover. In other words, the thicker the soil cover, the more transport-limited a process can be classified.

Gilbert's, Jahn's and Kirkby's definitions can be expanded to other fields in geomorphology. For example, a distinction was made between weathering-limited and transport-limited watersheds to explain DEBRIS FLOW activity (Bovis and Jakob 1999). In their paper, the authors define transport-limited WATERSHEDS as those in which there is a quasi-infinite amount of material available for transport by debris flows, while weathering-limited watersheds are those in which sediment supply rates to the channel system (by ravelling, rockfall, debris slides, etc.) are low and a channel requires recharge over a given time period before a rainstorm can trigger the next debris flow. This distinction is important in predicting debris flow frequencies as well as magnitudes. In the first watershed type (transport-limited) the exceedance of a climatic threshold will likely trigger a debris flow, while in the second watershed type (weathering-limited) a debris flow will only occur at the exceedance of a climatic threshold, if sufficient material has accumulated in the channel system. A similar separation has been made by Stiny (1910) in his monograph on debris flows written in German and later translated by Jakob and Skermer into English (1997). Stiny distinguished between 'Altschuttmuren' (old rubble debris flows) and 'Jungschuttmuren' (young rubble debris flows). The first are derived from large accumulations of debris (usually morainal material) with a quasi-infinite supply of material. The latter are derived from recently accumulated debris which may be depleted in the course of the debris flow.

Recently, the concept of weathering and transport limitation has been extended to encompass the development of fluvial BEDFORM. For example, during low flow stages of sand and GRAVEL-BED RIVERS, the channel bed may be armoured and sediment from upstream will pass over the armour layer in the form of sediment supply limited (weathering-limited) bedforms such as BARCHANS and sand ribbons (Kleinhans *et al.* 2002). Fully developed RIPPLES and dunes (see DUNE, FLUVIAL) exist when sufficient material exists for the formation of these features (transport-limited). The relevance of this distinction lies in the observation that bedform characteristics cannot be explained or predicted from local hydraulics and sediment characteristics alone. Another application is, for example, to assign controls on WEATHERING rates and solute (see SOLUTE LOAD AND RATING CURVE) fluxes on large rivers (Stallard 1985, 1995) using the same principles.

It is important to recognize that weathering-limited systems and transport-limited systems are not constant in space and time. For example, a large LANDSLIDE in a previously largely forested watershed may expose a large amount of debris which is now susceptible to transport by debris flows or other transport mechanisms. At this point the power of the stream (see STREAM POWER) network to transport the freshly accumulated and newly eroded material may have been exceeded and a previously weathering-limited watershed has now shifted to transport-limitation. Similarly, sudden climatic shifts may result in accelerated weathering rates which may not be matched by transport rates or vice versa. Finally, over time slope gradients may reach thresholds whereby the weathering-limitation may shift to transport-limitation.

The terms weathering and transport limited are somewhat misleading because they imply that the process of mechanical weathering is responsible for providing material to the system. However, other processes such as landsliding or resupply of material by fluvial processes may provide material which has not directly been supplied by weathering. Weathering is also hardly ever 'limited' compared to other geomorphic processes and is responsible for significant redistribution of material. Similarly, the term transport-limited may create confusion because it implies that it is the transport mechanism that limits the flux of mass. Rather than the transport mechanism, it is the availability of material that allows mass movement. It is therefore suggested to replace the terms weathering-limited and transport-limited by the less ambiguous terms 'supply-limited' and 'supply-unlimited'.

References

Bovis, M.J. and Jakob, M. (1999) The role of debris supply conditions in predicting debris flow activity, *Earth Surface Processes and Landforms* 24, 1,039–1,054.
Carson, M.A. and Kirkby, M.J. (1972) *Hillslope Form and Process*, Cambridge: Cambridge University Press.
Gilbert, G.K. (1877) *The Geology of the Henry Mountains*, Washington, DC: United States Geographical and Geological Survey.
Jahn, A. (1954). Denudational balance of slopes (in Polish), *Czasopismo Geograficzne* 25.
——(1968) Denudational balance of slopes, *Geographia Polonica* 13, 9–29.
Jakob, M. and Skermer, N. (translators) (1997) Stiny, J. *Die Muren*, Vancouver, BC: EBA Engineering Consultants Ltd.
Kleinhans, M.G., Wilbers, A.W.E., De Swaaf, A. and Van den Berg, J.H. (2002) Sediment supply-limited bedforms in sand-gravel bed rivers, *Journal of Sedimentary Research, Section A, Sedimentary Petrology and Processes* 72(5), 629–640.
Stallard, R.F. (1985) River chemistry, geology, geomorphology, and soils in the Amazon and Orinoco Basins, in J.I. Driver (ed.) *The Chemistry of Weathering*, 293–319, Dordrecht: Reidel.
——(1995) Tectonic, environmental, and human aspects of weathering and erosion: a global review using a steady-state perspective, *Annual Reviews of Earth and Planetary Science* 23, 11–39.
Stiny, J. (1910) *Die Muren. Versuch einer Monographie mit besonderer Berücksichtigung der Verhältnisse in den Tiroler Alpen*, Innsbruck: Verlag der Wagner'schen Universitäts Buchhandlung.

Further reading

Kirkby, M.J. (1971) Hillslope process-response models based on the continuity equation, *Transaction of the Institute of British Geographers*, Special Publication 3, 15–30.
——(1985) The basis for soil profile modelling in a geomorphic context, *Journal of Soil Science* 36, 97–121.

MATTHIAS JAKOB

WEATHERING PIT

Small closed depressions on horizontal and gently inclined rock surfaces. They are also called 'gnammas', 'Opferkessel' or 'pias'. They have been described from a range of silicate rock types, most frequently granites and sandstones. They may be similar in their broad morphology to solutional pits developed in carbonate rocks, to which the term 'kamenitza' is often applied. Goudie and Migoń (1997: Table 1) provides a list of references on weathering pits from a diverse range of morphoclimatic regions that range from polar to desert and humid tropical. The largest examples may be between 10 and 20 m long (Twidale and Corbin 1963).

There is a great deal of uncertainty concerning the processes involved in the development of pits. Chemical processes of solution are usually invoked, but other processes may include hydration, the mechanical action of frost and salt and biochemical weathering. Positive feedback mechanisms related to the ever-growing amount of water available as a pit enlarges may account for a localized high intensity of weathering (Schipull 1978).

Many weathering pits are reported to be partially infilled with debris and organic matter, yet there are also many which are empty, with bare flaky bedrock exposed on the floor. It seems that

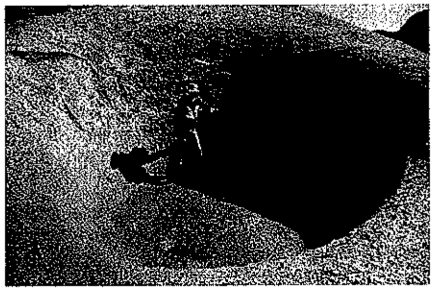


Plate 146 A large weathering pit developed in granite in the Erongo Mountains of central Namibia. This example has a completely bare bottom and the process by which debris has been removed from it is still the subject of debate

relatively little attention has been paid to the question of how the debris gets evacuated from the pit. Three possible ways have been suggested (Smith 1941), namely solutional transport and washing out during excessive rainfalls, and deflation. Flotation is another possible mechanism, but no direct observations have been made. Moreover, the occurrence of deep closed pits devoid of any sediments (cf. Watson and Pye 1985) remains puzzling and no satisfactory explanation of their emptiness has been offered.

References

Goudie, A.S. and Migoń, P. (1997) Weathering pits in the Spitzkoppe area, Central Namib Desert, *Zeitschrift für Geomorphologie* NF 41, 417–444.
Schipull, K. (1978) Waterpockets (Opferkessel) in Sandsteinen des zentralen Colorado-Plateaus, *Zeitschrift für Geomorphologie* NF 22, 426–438.
Smith, L.L. (1941) Weathering pits in granite of the Southern Piedmont, *Journal of Geomorphology* 4, 117–127.
Twidale, C.R. and Corbin, E. (1963) Gnammas, *Revue de Géomorphologie Dynamique* 14, 1–20.
Watson, A. and Pye, K. (1985) Pseudokarstic microrelief and other weathering features on the Mswati granite (Swaziland), *Zeitschrift für Geomorphologie* NF 29, 285–300.

A.S. GOUDIE

WETTING AND DRYING WEATHERING

Fluctuations in rock moisture – wetting and drying – can cause rock weathering. In many cases these fluctuations cause weathering as a result of the rock expanding during take up of water and its inability to return to the original dimensions upon losing moisture. Moreover, high moisture contents can diminish rock strength, and through time wetting and drying cycles can reduce the bonding strength of component minerals. This in turn leads to a decrease in rock strength and possibly even failure (Pissart and Lautridou 1984). Cycles of wetting and drying can be frequent in some environments (e.g. shore platforms) and experimental simulations have shown the effectiveness of the process (Hall and Hall 1996). Micro-erosion metre studies have demonstrated the importance of surface swelling on shore platforms developed on limestones and mudstones, though it is less easy to discriminate between the importance of the growth of salt crystals and the expansion from wetting and drying (Stephenson and Kirk 2001). Indeed, wetting and drying is frequently a sine qua non for salt weathering. Wetting and drying also operates on badland surfaces developed on mudstones in semi-arid areas (Cantón *et al.* 2001), though once again it is the combined effect of moisture cycling and salt dissolution that is crucial. Nonetheless, clay-rich rocks are susceptible to SLAKING with or without the presence of salts (Gökçeuğlu *et al.* 2000), and the process can cause severe disintegration, as with the tombs in the Valley of the Kings, Egypt, which have been excavated in Esna Shales (Wüst and McLane 2000).

References

Cantón, Y., Solé-Benet, A., Queralt, I. and Pini, R. (2001) Weathering of a gypsum-calcareous mudstone under semi-arid environment at Tabernas, SE Spain: laboratory and field-based experimental approaches, *Catena* 44, 111–132.
Gökçeuğlu, C., Ulusay, R. and Sönmez, H. (2000) Factors affecting the durability of selected weak and clay-bearing rocks from Turkey, with particular emphasis on the influence of the number of drying and wetting cycles, *Engineering Geology* 57, 215–237.
Hall, K. and Hall, A. (1996) Weathering by wetting and drying: some experimental results, *Earth Surface Processes and Landforms* 21, 365–376.
Pissart, A. and Lautridou, J.P. (1984) Variations de longeur de cylinders de Pierre de Caen (calcaire bathonien) sous l'effet de séchage et d'humidification, *Zeitschrift für Geomorphologie NF* 29, 111–116.
Stephenson, W.J. and Kirk, R.M. (2001) Surface swelling of coastal bedrock on inter-tidal shore platforms, Kaikoura peninsula, South Island, New Zealand, *Geomorphology* 41, 5–21.
Wüst, R.A.J. and McLane, J. (2000) Rock deterioration in the Royal Tomb of Seti I, Valley of the Kings, Luxor, Egypt, *Engineering Geology* 58, 163–190.

A.S. GOUDIE

WILSON CYCLE

The Wilson cycle is the hypothesis that oceans are born as rifts, grow by SEAFLOOR SPREADING, and finally close again (Wilson 1966). Six stages are identified:

1 Uplift and extension forming rift valleys (see RIFT VALLEY AND RIFTING) (like modern rift valleys in Africa).
2 Early seafloor spreading (Red Sea stage).
3 Mature ocean (Atlantic stage) with a broad ocean bounded by continental shelves and a spreading centre at the mid-ocean ridge.
4 Shrinking of the ocean, bounded by ISLAND ARCS. The Pacific is in this stage, though the Pacific is also spreading.
5 Further shrinking, compression, metamorphism and uplift of accretionary wedges to form mountains (Mediterranean stage).
6 All oceanic crust is subducted, and the continents converge on a collision-zone suture (e.g. Indus–Tsangpo suture in the Himalayas). The united plate breaks along a new rift, restarting the cycle.

Some think the cycle should be repeated by rupture along roughly the same old line of weakness, others think a new cycle can start along a new fracture. Van der Pluijm and Marshak (1997) believe that rifting may occur in warm and weak orogens. Since orogenic belts are regarded as the weaker portions of the crust, they tend to be sites of repeated rifting and collision. Rocks in a single orogen may record the effects of several phases of extension and contraction. In the eastern United States, for example, the Earth history involves two phases of rifting (Late Precambrian and Middle Mesozoic) and two phases of collision (Late Precambrian and Palaeozoic), with several subphases (van der Pluijm and Marshak 1997).

The apparent polar wander (APW) path of two continents experiencing a Wilson cycle will show parallel tracks that diverge at breakup and converge after closure to a second parallel track. Such palaeomagnetic support for the Wilson cycle has been claimed for the rifting and closure of the Atlantic (Piper 1987). Gondwanaland and Laurasia converged during Silurian times to form Pangaea. From then to the Mesozoic they had a common APW path, showing they formed a single continent. They then split up into the present continents along new lines, continuing the Wilson cycle.

The opening and closing of ocean basins up to 1,500 km wide takes about 500 Ma, but some are smaller and more short-lived. Because of the timescale, evidence for Wilson cycles comes mainly from plate tectonic interpretation of palaeomagnetic, structural and geological data. In general it has little direct effect on geomorphology, but Russo and Silver (1996) relate the formation of the Andes to the Wilson cycle.

References

Piper, J.D.A. (1987) *Palaeomagnetism and the Continental Crust*, Milton Keynes: Open University Press.
Russo, R.M. and Silver, P.G. (1996) Cordillera formation, mantle dynamics, and the Wilson Cycle, *Geology* 24, 511–514.
Van der Pluijm, B.A. and Marshak, S. (1997) *Earth Structure: An Introduction to Structural Geology and Tectonics*, New York: WCB/McGraw-Hill.
Wilson, J.T. (1966) Did the Atlantic close and then re-open? *Nature* 211, 676–681.

CLIFF OLLIER

WIND EROSION OF SOIL

Wind erosion (see AEOLATION) has for long been recognized as an important factor in the development of landforms in dryland regions, contributing to the development of such features as close depressions (PANS), streamlined hillocks (YARDANGS), deflated surfaces (STONE PAVEMENTS) and miscellaneous micro-forms (VENTIFACTS). Wind erosion may also modify the form of dunes, as, for example, by creating blowouts (see DUNE, COASTAL). The general context of wind erosion is discussed in various other entries (see AEOLIAN GEOMORPHOLOGY; AEOLIAN PROCESSES; DESERT GEOMORPHOLOGY, etc.). This entry concentrates on the erosion of soil by wind – one component of DESERTIFICATION.

Wind erosion is a natural erosion process. However, its intensity and magnitude is often increased by miscellaneous human activities such as cultivation, overgrazing, vehicular traffic, etc.

(Leys 1999). It has negative impacts on agricultural production and produces environmental pollution, including DUST STORMS. Attempts to understand and model wind erosion have been made for some decades, most notably by W.S. Chepil and colleagues from the United States Department of Agriculture (see Chepil and Woodruff 1963). Using empirical field studies and wind tunnels (see WIND TUNNELS IN GEOMORPHOLOGY), they developed a much used Wind Erosion Equation.

$$E = f(C, I, L, K, V)$$

where E is the potential loss, C is a local climatic index, I is a soil erodibility index, L is a factor relating to field shape in the prevailing wind direction, K is a ridge roughness factor for ploughed ground and V is a vegetation cover index. The equation was developed initially in the Midwest of the USA and drew attention to the factors which could be manipulated by farmers (namely, I, L, K and V).

The climatic factor (C) was a simple combination of two key climatic variables: annual wind speed and a moisture index. Plainly, dry, windy areas are likely to be most susceptible to wind erosion. The soil erodibility factor (I) is more complex and needs to be seen in terms of both individual grain size characteristics and aggregate characteristics. Fine sands and silts are likely to be most susceptible, partly because of the relatively low velocities required for their entrainment, but also because the presence of clay tends to produce wind-stable clods. The presence of large clods reduces the risk of wind erosion. The fetch distance over which the wind acts (L) is related to field size and the presence or absence of shelter belts of differing heights, spacing and permeability. The ridge roughness factor (K) is based on the experimental observation that the rougher the surface, up to about 6 cm, the lower the windspeed at the surface. Thus furrows at right angles to the wind will tend to dampen down rates of erosion. The vegetation factor (V) is absolutely fundamental, for a dense vegetation cover, especially if like grass it has short stalks and narrow leaves, does more than anything else to reduce erosion rates.

More recently predictive models of wind erosion have been developed (Shao 2000). One such model is the Wind Erosion Assessment Model (WEAM), which aims to account for the combined effect of climate, soil, vegetation and land use. The fundamental physical viewpoint of this model (Shao *et al.* 1996) is that wind erosion is a result of two opposing forces: the capacity of the wind to start and maintain erosion, and the ability of the soil to resist it. The wind's capacity to start and maintain erosion is the friction velocity u^* (the wind shear or drag on the soil), while the opposing quantity offered by the soil is the threshold velocity u^*_t (the minimum friction velocity that is required for erosion to occur). The former is determined by wind flow conditions and the surface roughness, while the latter is determined by such surface factors as soil texture, aggregate structure and moisture content.

References

Chepil, W.S. and Woodruff, N.P. (1963) The physics of wind erosion and its control, *Advances in Agronomy* 15, 211–302.

Leys, J. (1999) Wind erosion on agricultural land, in A.S. Goudie, I. Livingstone and S. Stokes (ed.) *Aeolian Environments, Sediments and Landforms*, 143–166, Chichester: Wiley.

Shao, Y. (2000) *Physics and Modelling of Wind Erosion*, Dordrecht: Kluwer.

Shao, Y., Raupach, M.R. and Leys, J.F. (1996) A model for predicting aeolian sand drift and dust entrainment on scales from paddock to region, *Australian Journal of Soil Research* 34, 309–342.

A.S. GOUDIE

WIND TUNNELS IN GEOMORPHOLOGY

Wind tunnels provide a means by which processes in AEOLIAN GEOMORPHOLOGY can be monitored with a fully controlled and regulated wind flow. This is particularly advantageous in the aeolian environment where natural fluctuations in wind velocity and direction make it difficult to perform repeatable experiments in the field.

Wind tunnels vary greatly in design characteristics. They usually consist of a fan sucking (or blowing) air through a contraction which flattens out streamlines and provides a non-fluctuating wind to a working section where experiments can be carried out. The size of the working section may vary from $0.1\,m^2$ and 1 m in length to more than $4\,m^2$ and over 10 m in length. Working velocities in wind tunnels are commonly between 1 and $20\,m\,s^{-1}$.

Geomorphological experiments in wind tunnels are normally concerned either with aeolian sand or dust transport processes (Butterfield 1998; see AEOLIAN PROCESSES) where sediment is placed in the working section, or reduced scale studies in 'clean' wind tunnels of flow over fixed models of dunes, hills or valleys (Walker and Nickling 2002). In the latter case not only must the landform be reduced in scale but also the structural characteristics of the windflow (the atmospheric boundary layer). This is often achieved through a series of upstream grids and spires which, together with a roughened working section floor, can provide scaled values of shear velocity, aerodynamic roughness and turbulence length scales which mimic full-scale values. Portable wind tunnels are commonly used to test wind erosion (SEE WIND EROSION OF SOIL) characteristics of agricultural fields. Here the floor of the tunnel working section is absent so that when placed at a test site the tunnel flow acts directly on the natural soil.

Wind flow in wind tunnels can be determined with a pitot-tube which establishes velocity through the measurement of pressure. Where highly turbulent flows are encountered then electronic hot-wire anemometers are more commonly employed. Precise measurement of shear stresses on the surface of scaled models can now be accomplished using pulse-wire anemometers (Wiggs *et al.* 1996) or particle velocimetry whilst flow visualization over scale models is achievable by seeding the flow with smoke or covering the surface of the model with oil. In experiments involving sediment entrainment and transport sand traps can be erected at the downwind end of the working section to measure the rate of sand flux.

Whilst many advances have transpired as a result of the use of wind tunnels in research on aeolian processes and wind flow over aeolian bedforms, the biggest drawback of the approach is that there is commonly insufficient field-based empirical data with which results can be verified.

References

Butterfield, G.R. (1998) Transitional behaviour of saltation: wind tunnel observations of unsteady winds, *Journal of Arid Environments* 39, 377–394.

Walker, I.J. and Nickling, W.G. (2002) Dynamics of secondary airflow and sediment transport over and in the lee of transverse dunes, *Progress in Physical Geography* 26, 47–75.

Wiggs, G.F.S., Livingstone, I. and Warren, A. (1996) The role of streamline curvature in sand dune dynamics: evidence from field and wind tunnel measurements, *Geomorphology* 17, 29–46.

GILES F.S. WIGGS

Y

YARDANG

A Turkmen word to describe wind abraded ridges of cohesive material (Hedin 1903). They have been described from a large number of arid regions (McCauley *et al.* 1977) where they have developed on a large range of materials. They have been likened in shape to the prows of upturned ships. They range in size from small centimetre-scale ridges (micro-yardangs) through to forms that are some metres in height and length (meso-yardangs), to features that may be tens of metres high and some kilometres long (mega-yardangs). The forms may go through a cycle of development and eventual obliteration (Halimov and Fezer 1989). Although they are dominantly aeolian erosion features there has been a considerable debate as to the relative importance of deflation, aeolian abrasion, fluvial incision and mass movements in moulding yardang morphology. That abrasion is important is indicated by polished, fluted and sand-blasted slopes, and the undercutting of the steep windward face and lateral slopes. It is probably the dominant process in hard bedrock yardangs whereas deflation may be important in the evolution of yardangs developed in soft sediments such as old lake beds. Fluvial erosion may provide an avenue along which wind erosion may occur but excessive fluvial erosion would tend to obliterate yardangs. Mass movements may also be significant when yardang slopes have been oversteepened by wind erosion.

One form of yardang, beloved of textbooks, but in reality not very common or significant, is the zeuge (plural zeugen). The term is derived from the German word for 'witness'. Zeugen are tabular masses of resistant rock (2–50 m high) left standing on a pedestal of softer material as a result of differential erosion by sand-laden wind.

Plate 147 Characteristic aerodynamic yardangs developed by the wind erosion of Holocene pluvial lake beds in the Dhakla Oasis, Western Desert, Egypt

Plate 148 A series of wind eroded chalk outcrops in the White Desert, Farafra, Egypt. In addition to wind erosion, note the presence of case hardening and honeycomb weathering

At the mega-scale yardangs curve round in response to the trajectories of trade winds (as, for example, around Borkou in the Central Sahara; Mainguet 1972) and may show a remarkable parallelism over hundreds of kilometres (as with the *Kaluts* of the Lut Desert in Iran).

Study of the morphometry of yardangs has revealed relationships between different parameters. Ward and Greeley (1984) found a 1:4 width to length ratio; Halimov and Fezer (1989) found that the ratios of length, width and height were 10:2:1; while Goudie *et al.* (1999) found volume, length, width, height ratios of 18.7:9.9:2.7:1.

In soft materials yardangs can form quickly. In the Sahara, Mojave and Lop Nor deserts, Holocene lake and swamp deposits have been eroded at rates of *c.*2.5 to 5 m per 1,000 years.

References

Goudie, A.S., Stokes, S., Cook, J. *et al.* (1999) Yardang landforms from Kharga Oasis, south-west Egypt, *Zeitschrift für Geomorphologie Supplementband* 116, 1–16.

Halimov, M. and Fezer, F. (1989) Eight yardang types in Central Asia, *Zeitschrift für Geomorphologie* 33, 205–217.

Hedin, S. (1903) *Central Asia and Tibet*, Scribners: New York.

McCauley, J.F., Grolier, M.J. and Breed, C.S. (1977) Yardangs of Peru and other desert regions, *US Geological Survey Interagency Report: Astrogeology* 81.

Mainguet, M. (1972) *Le Modelé des Grès*, Paris: IGN.

Ward, A.W. and Greeley, R. (1984) Evolution of yardangs at Rogers Lake, California, *Geological Society of America Bulletin* 95, 829–837.

A.S. GOUDIE

YAZOO

A tributary stream/river that flows for a long distance parallel to the main channel before joining it. This usually occurs in mature river channels where natural levees have built up on the channel banks, above the mean height of the surrounding floodplain. The tributary channel is therefore unable to penetrate the trunk channel and lakes may form. However, more often the tributary channel is bounded by bluffs beyond the floodplain, thus forcing a Yazoo-type tributary downstream parallel to the main channel. The type example is the Yazoo River, which runs south/south-west parallel to the main Mississippi River from Memphis, Tennessee (USA) for almost 400 km. The Yazoo eventually infiltrates the Mississippi channel at Vicksburg, Mississippi. The elevated bed of the Mississippi River obstructs the Yazoo tributary and forces it to flow parallel along the floodplain (35–150 km wide) and with a Yazoo channel gradient of only 11.4 cm per km. Yazoo-type channels are also evident in deep sea environments (Hesse and Rakofsky 1992).

Reference

Hesse, R. and Rakofsky, A. (1992) Deep sea channel submarine yazoo system of the Labrador Sea – a new deep-water facies model, *American Association of Petroleum Geologists Bulletin* 75(5), 680–707.

Further reading

Dorris, F.E. (1929) The Yazoo Basin in Mississippi, *Journal of Geography* 28, 72–81.

STEVE WARD

Z

ZOOGEOMORPHOLOGY

Zoogeomorphology is the study of the geomorphic effects of animals (Butler 1995), and as such can be considered as a subset of BIOGEOMORPHOLOGY. Zoogeomorphology encompasses the geomorphic effects of both free-ranging, wild animal populations as well as the geomorphic effects of domesticated animals such as cattle. It considers the geomorphic role of animals regardless of their size, ranging from small invertebrates such as ants, worms and termites (see TERMITES AND TERMITARIA) to both terrestrial and aquatic vertebrates including fish, amphibians, reptiles, birds and mammals.

Animals geomorphologically alter the Earth's surface through both direct and indirect effects. Direct effects encompass surface erosion, transportation and deposition of rock, soil and/or unconsolidated sediments by animals. Animals accomplish these effects by digging for and/or catching of food, nest building, burrowing, mound-building, wallowing, eating soil or rocks, canal excavation and dam building (accomplished by beavers; *Castor canadensis* in North America, *C. fiber* in Eurasia). Indirect effects do not specifically remove sediment from a surface, but processes such as trampling associated with overgrazing may lead to a reduced infiltration capacity in the underlying soil, in turn resulting in greater surface wash, soil creep and rainsplash detachment. Trampling and associated overgrazing by domesticated animals also contribute to widespread bank erosion, riparian degradation and unstable channel morphologies (Trimble 1994; Magilligan and McDowell 1997; Naiman and Rogers 1997; Evans 1998).

The geomorphic impacts of animals may vary spatially and temporally through the course of a year, as wild animals migrate in search of changing food sources (Baer and Butler 2000). In the Rocky Mountains of western North America, for example, grizzly bears (*Ursus arctos horribilis*) create widespread diggings during the spring on lower floodplains in search of roots and tubers. As summer progresses, they migrate upslope into areas of snowmelt where they dig for burrowing mammals such as ground squirrels as well as for roots. Late summer sees the bears on high talus slopes, where they excavate large quantities of boulders, in search of insect larvae.

The seasonality of geomorphic impacts by animals is also a function of a region's climate. Excessive heat and dryness minimize the summertime geomorphic impacts of some desert dwellers. In colder climates, frozen ground limits the seasonal ability of animals to dig into the surface. During such periods, many burrowing animals may also hibernate, effectively cutting off geomorphic impacts until the subsequent spring (Butler 1995).

Beavers have some of the most profound geomorphic impacts of any animal in North America. Beavers are large (adult mass >15 kg), semi-aquatic rodents that instinctively attempt to build dams in response to the sound of running water (Butler 1995). Prior to European settlement of North America, as many as 400 million beavers may have occupied an area of approximately fifteen million square kilometres. By the year 1900, trapping of beavers in North America had nearly extinguished the species. During the twentieth century, conservation efforts re-established the beaver throughout their native range, but at population levels of only 10 per cent of the pre-European-contact level. As beavers continue to reoccupy their former range, their geomorphic influences also grow. For example, on the Roanoke River floodplain, on the Coastal Plain of the eastern United States, beaver ponds illustrated a tenfold increase in areal coverage between the years 1984 and 1993 (Townsend and Butler 1996).

Beaver dams impound water, creating pond environments into which streams flow and deposit sediment. Individual beaver ponds may accumulate as much as 20–30 cm of sediment across the floor of a pond annually (but only during the portions of the year when the inflowing stream is not frozen). Older ponds entrap significantly more sediment than do younger ponds in topographically similar sites (Butler and Malanson 1995). Beaver dams substantially reduce stream velocity and discharge, to the point in some cases of completely eliminating surface outflow on the downstream side of a dam (Meentemeyer and Butler 1999).

Animal burrows (such as wombats in Australia) and mounds (termites) may cover as much as several cubic metres, and are visible on aerial photographs and even satellite imagery (Löffler and Margules 1980). Some of the most widespread geomorphic impacts of animals are produced on the floor of the Bering Sea by the California gray whale (*Eschrichtius robustus*). Gray whales feed almost exclusively on bottom-dwelling organisms, especially amphipod crustaceans. The whales 'slurp' deep furrows and pits into the floor of the ocean, ingesting sediment and fauna. The whales filter the food from the sediment in their baleen, and subsequently expel plumes of muddy sediment near the surface (Butler 1995). These plumes, visible from low-flying airplanes, account for an enormous amount of sediment transport in the Bering Sea. Annual whale feeding there moves a minimum of 120 million cubic metres of sediment, nearly three times the annual load of suspended sediment discharged into the Bering Sea by the Yukon River, the fourth largest sediment source in North America.

References

Baer, L.D. and Butler, D.R. (2000) Space-time modeling of grizzly bears, *Geographical Review* 90, 206–221.

Butler, D.R. (1995) *Zoogeomorphology – Animals as Geomorphic Agents*, Cambridge and New York: Cambridge University Press.

Butler, D.R. and Malanson, G.P. (1995) Sedimentation rates and patterns in beaver ponds in a mountain environment, *Geomorphology* 13, 255–269.

Evans, R. (1998) The erosional impacts of grazing animals, *Progress in Physical Geography* 22, 251–268.

Löffler, E. and Margules, C. (1980) Wombats detected from space, *Remote Sensing of Environment* 9, 47–56.

Magilligan, F.J. and McDowell, P.F. (1997) Stream channel adjustments following elimination of cattle grazing, *Journal of the American Water Resources Association* 33, 867–878.

Meentemeyer, R.K. and Butler, D.R. (1999) Hydrogeomorphic effects of beaver dams in Glacier National Park, Montana, *Physical Geography* 20, 436–446.

Naiman, R.J. and Rogers, K.H. (1997) Large animals and system-level characteristics in river corridors, *BioScience* 47, 521–529.

Townsend, P.A. and Butler, D.R. (1996) Patterns of landscape use by beaver on the lower Roanoke River floodplain, North Carolina, *Physical Geography* 17, 253–269.

Trimble, S.W. (1994) Erosional effects of cattle on streambanks in Tennessee, U.S.A., *Earth Surface Processes and Landforms* 19, 451–464.

Further reading

Butler, D.R. (1995) *Zoogeomorphology – Animals as Geomorphic Agents*, Cambridge and New York: Cambridge University Press.

SEE ALSO: biogeomorphology; termites and termitaria

DAVID R. BUTLER