ICE

Ice is the solid phase of water, i.e. a chemical compound of two positive univalent ions of hydrogen and one negative bivalent oxygen (H_2O). Water and ice are responsible for most of the processes leading to landscape sculpturing. In water an isosceles triangle is formed by the three nuclei of the H_2O and the apex angle is equal to 105°3′. The oxygen nucleus is at the apex and the hydrogen nuclei (protons) at the other two corners of the molecule. Two of the eight electrons of the oxygen atom rotate close to the nucleus and another two have eccentric orbits which also contain the electrons from one of the hydrogen atoms. The oxygen nucleus and one proton are enclosed by each of these orbits. The other four electrons circle in two other eccentric orbits. For this reason, four eccentric orbits radiate tetrahedrally from the oxygen nucleus, which is completely screened by the electron orbits. Some of the positive charge of the protons is not screened completely. The eccentric electron orbits provide an excess negative charge in the direction of the two orbits without protons. Small negative charges at the oxygen end of the molecule and equally small positive charges at the hydrogen end determine that the H_2O molecules are slightly charged. Thus, a weak electrostatic bond is formed between the hydrogen end of one molecule and the oxygen end of another (the hydrogen bond) and it is for this reason that each molecule of water is surrounded by four others in a regular tetrahedral spatial arrangement. With respect to any particular molecule, the hydrogen bond is more than ten times weaker than the covalent bond. Therefore, mixtures of molecules are easily formed and broken by the reduction or addition of energy. In the liquid phase, the molecules are in motion and can be of variable distances, one to another, but, in ice, stable hydrogen bonds are created between the surrounding molecules, thereby generating the hexagonal crystal structure (Figure 90a, b). This is because in negative temperatures the energy of the system is lower. It gives pure ice a symmetrical lattice arrangement and a lower density (0.91668 g cm^{-3}) than pure water (0.999984 g cm^{-3} – at the temperature 0 °C and the normal atmospheric pressure 1013 hPa). The molecules within the ice crystal are organized in layers of hexagonal rings. The atoms in a ring are not in one plane but two (Figure 90b). Spacing between two layers (0.276 nm) is much larger than between two planes of atoms (0.0923 nm). Adjacent layers form mirror images of each other. The optic axis (also called the *c-axis* or *principal axis*) of the ice crystal is perpendicular to the *basal plane*, i.e. the plane of a layer of hexagonal rings of atoms. Within the hexagonal lattice, three crystallographic *a-axes*, separated by 60° from each other, form the basal plane (Figure 90a, c).

A dozen crystallographic kinds of ice are known in nature and laboratory experiments but only two are observed in the Earth's natural conditions. The most common is the hexagonal crystalline ice which exists in temperatures between 0 °C and *c.* −70 °C and pressures up to *c.*210 MPa. Ice in the cubic crystalline form has been found at very low temperatures (below *c.* −70 °C) and very low pressures in the upper parts of the troposphere. Other ice polymorphs can appear in very high pressures (higher than 800 MPa) and at various positive temperatures. Such 'hot ices' can reach density higher than 1.3 g cm^{-3} and might be present in the lithosphere as films of solid water, only a few molecules thick, chemically bonded

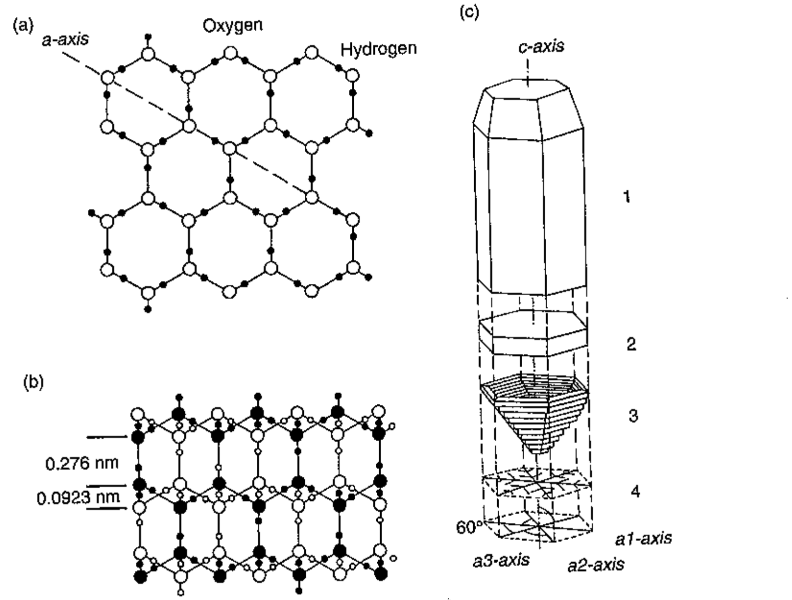

Figure 90 Structure of ice and forms of crystals (modified from Jania 1988). Models of the ice lattice structure: (a) basal plane, (b) view perpendicular to the c-axis. Types of the hexagonal ice crystals (c): 1 – column (prism), 2 – plate, 3 – cup, 4 – snow star

with other minerals. It is thought that these may affect certain properties of rocks, e.g. their susceptibility to weathering after exposure at the land surface.

The increase of volume of freezing water, by *c.*10 per cent, is responsible for important geomorphic processes (e.g. MECHANICAL WEATHERING; SOLIFLUCTION). The higher specific volume of the 'common' hexagonal ice compared with water, especially in respect of the relatively more dense sea water, enables ice to flow. Floating ice can be also a geomorphic agent (e.g. ICEBERGS). *Drift ice deposits* resulting from melting of debris-rich icebergs are recorded not only in deep sea deposits but also in shallow bays and proglacial lakes.

Six isotopes of the oxygen and three isotopes of hydrogen have been noted in natural conditions. Their various combinations result in thirty-six isotopic kinds of water and ice. The most important are: $H_2{}^{16}O$ (common water); $HD^{16}O$ ('heavy water' with deuter – D) and $H_2{}^{18}O$ (water with 'heavy' oxygen – ^{18}O). The last two are minor admixtures (0.03 per cent and 0.2 per cent, respectively) of ocean waters which are the most important source of water vapour in the atmosphere and consequently of snow and ice on land. Molecules of water with the heavier stable isotopes evaporate less rapidly and condense slightly faster from the vapour. The isotopic composition of snow precipitation depends on the temperature of evaporation of the sea water and the temperature of condensation (re-sublimation). The proportion of contents of heavy isotopes is a fraction of their concentration relative to the 'light' isotopes ($^{18}O/^{16}O$ or D/H) in a sample of snow (ice). In respect of its deviation (δ) from the ratio in the 'standard mean ocean water' (SMOW) it is expressed in parts per thousand (‰). The average annual values of $\delta^{18}O$ in snow samples in the polar areas correlates closely (in linear regression) with mean annual air temperatures there. It follows that the stable isotope content in samples from deep ice cores is a good source of data on air temperatures in the past (paleotemperatures).

Despite the fact that the hexagonal crystalline ice (Figure 90c) is the most common on the Earth, perfectly symmetrical hexagonal ice crystals are in the minority in the natural environment. Varied factors, which influence crystallization of water, determine that ice crystals are often incomplete or defective. First, natural waters are very rarely chemically pure. They have dissolved admixtures which lead to the phase change (crystallization) at temperatures lower than 0.16°C – the figure usually regarded as the freezing temperature. Water in a supercooled state (i.e. being in the liquid form below the freezing point temperature) rarely freezes spontaneously. It is able to freeze only when crystallization centres are available within the liquid (e.g. small particles of ice or other minerals). When water freezes, H_2O molecules arrange themselves in a lattice of crystals. When cooling is slow in still waters, other molecules are removed to outside the walls of the ice crystals. Fast crystallization and turbulent movement of water cause incorporation of molecules of other chemical species into the lattice of ice crystals and their structure becomes defective. In general, optic axes of ice crystals are oriented perpendicularly to the 'cooling front' or the freezing surface. Vertical ice crystals are characteristic of lake ice and *naled ice*, whereas *rime ice* may have c-axes oriented toward the wind direction. *Glacier ice* and the anchor and frazil river ice usually have randomly oriented crystal optic axes.

Defects within the crystal lattice are regarded as factor facilitated recrystallization and ice crystals become reoriented when new directions of shear stress are applied. Under high cryostatic pressure, atmospheric gas molecules (e.g. N_2, O_2, CO_2) from air bubbles, trapped between glacier ice crystals, are incorporated into their lattice in the form of clatrates. When such ice crystals appear on the surface or in front of a glacier, where pressure is lower, these gas molecules are released from the lattice to either meltwater or the atmosphere.

Ice under natural conditions may variously be regarded as a mineral, sediment or a rock (Shumskyi 1955: 15–16). The hardness of ice varies with temperature. At 0°C, its hardness is 1.5 on the Mohs' scale (as is talc and gypsum), whereas at −40°C, the ice hardness is as much as 4 (equivalent of the hardness of fluorite). It is clear that glacier erosion would be impossible without the regelation process at the glacier bed, i.e. melting (under pressure) at the proximal side of subglacial obstacles and refreezing in their lee side.

Ice is common on the planet Earth and is also present on Mars, Pluto and satellites of Jupiter (e.g. Europa, Callisto) and Saturn. However, the presence of water in vapour, liquid and solid forms makes the Earth a planet unique to the Solar System. An irregular envelope containing ice encircles the Earth and is called the cryosphere. The term, introduced by Dobrowolski (1923), derives from the Greek (*kryos* – cold). The cryosphere exists only in an intermingled state with the lithosphere, hydrosphere and atmosphere, where water appears in the form of ice (snow cover, glaciers, sea and fresh waters ice, ground ice and ice in the atmosphere). The total mass of all ice on the Earth is estimated at 2.5×10^{16} metric tons. This is present on the surface or in the near-surface layers of the crust of our planet and occupies an area of *c.* $73.4 \times 10^6\,km^2$. It covers 14.2 per cent of the total area of the globe with annual fluctuations between 10.5 and 17.9 per cent. Ice, therefore, can hardly be other than a vastly important geomorphic agent, especially in polar and temperate climates. Considering all forms of natural ice, GLACIERs are the most important agents of relief changes. Total volume of glaciers and ice sheets constitutes about 97.7 per cent of all natural forms of ice on Earth, while subsurface ice is only 2.1 per cent (Kotlyakov 1984: 347–348).

Atmospheric ice, in the form of stars, plates or columns (Figure 90c), forms a snow cover when deposited on the ground. This mixture of ice crystals and their inclusions of air and water (in warmer environments) normally has a density between 0.05 and $0.5\,g\,cm^{-3}$. A seasonal snow cover affects slope processes, including rapid and spectacular mass movement as avalanches (see AVALANCHE, SNOW).

In areas where accumulation of snow exceeds its melting, snow is transformed into glacier ice. The chain of processes leading to the metamorphism of a fresh snow cover into this specific form of polycrystalline 'ice rock' differs in warm (temperate) environments (with presence of water within the snow cover) and in dry cold ones. Infiltration of meltwater into the snowpack accelerates the transformation. In cold conditions sublimation plays an important role in the rounding of crystals and their growth. Despite the environmental differences, four steps of transformation of snow into polycrystalline glacier ice are usually distinguished: the first is from fresh snow (porosity *c.* 95 per cent) to old snow (density 0.3–$0.5\,g\,cm^{-3}$) due to setting, compaction and rounding of crystals (step 1); further densification and compaction by pressure of overlying layers leads (step 2) from old snow (porosity *c.*50 per cent) into coarse-grained snow called *firn* (density $0.55\,g\,cm^{-3}$) about a year later. Firn has a low porosity, but is still permeable. Transformation from firn to the white (new) glacier ice (step 4) takes a longer time in cold climates, where intergranular movements (due to compaction), combined with sintering (recrystallization by way of sublimation) and plastic deformation, dominate. In warmer climates, owing to liquid water infiltration, metamorphosis is faster; this is due to refreezing and, of course, settling. In extremely cold and dry conditions, such processes takes hundreds or thousands of years (*c.*4,000 years at the Vostok Station, East Antarctica). In temperate and subpolar climates, it lasts decades. New glacier ice has a density higher than $0.83\,g\,cm^{-3}$, and, owing to closure of the pores into air bubbles between ice crystals, very low permeability. At this stage, crystals usually have dimensions of a couple of millimetres and their c-axes are randomly oriented. The last step in the growth of crystals and densification of glacier ice occurs through dynamic metamorphism and requires the deformations caused by *glacier flow* (see GLACIER). The old glacier ice may have densities of as much as $0.9\,g\,cm^{-3}$ and crystal sizes can have diameters larger than 10 cm (Plate 63). The optic axes of dynamically transformed glacier ice crystals reveal a predominant orientation, which is perpendicular to the terminal stress field. However, the *basal ice* of glaciers possesses a different structure. Most glaciers have an irregular bed with resistant rock protuberances. Melting develops when the sliding sole of a glacier passes such an obstacle in its bedrock (in thermal conditions close to the pressure melting point of ice). On the lee side of the obstacle, water refreezes, forming a layer of REGELATION ice and basal debris can be incorporated into this. The formation of debris-rich basal ice is observed in many temperate and polythermal glaciers. The isotopic composition of basal ice layers suggests an accretion of ice crystals from supercooled water from the subglacial drainage (Titus *et al.* 1999: 43). The debris content incorporated into basal ice (a basal moraine) is different, is very variable and depends on a number of factors. The basal debris zone tends to be thicker in the cold-based ice sheets of Greenland and Antarctic (up to16 m) than in temperate or polythermal glaciers, where its thickness varies from 0.4 m to a couple of metres. Debris

concentration in the thick dispersed basal debris zones reaches 7–12 per cent (by volume), whereas in the thin basal zones (<1 m) of warm-based glaciers, the concentration can exceed 50 per cent (Menzies 2002: table 6.2). In contrast, the englacial concentration of debris is generally very low (<1 per cent). Glacier sediment discharge depends on the debris concentration in the basal ice zone and the basal sliding velocity.

An ICE SHEET is a shield-like, broad ice mass of continent scale (larger than $50{,}000\,km^2$). It can be thick enough (thousands of metres) to cover any irregularity in topography of its bed completely. Ice flow is generally organized in a quasi-radial pattern from one or more centres: *ice domes*. The Antarctic Ice Sheet is the contemporary example ice mass of such dimensions: *c.* $13 \times 10^6\,km^2$, with volume of $30.11 \times 10^6\,km^3$. The ice cover of Antarctica has five ice domes which reach an elevation of 4,000 m a.s.l. The maximum thickness of ice (4,776 m) has been found in East Antarctica and its mean thickness is 2,160 m (Drewry 1983: 4). The Greenland Ice Sheet covers *c.* $1.75 \times 10^6\,km^2$ and has a mean ice thickness of 1,790 m (volume: $2.74 \times 10^6\,km^3$). Two well-pronounced ice domes are distinguished in Greenland and the northern dome reaches an elevation of 3,236 m a.s.l. (Van der Ween 1999). An ice dome has a

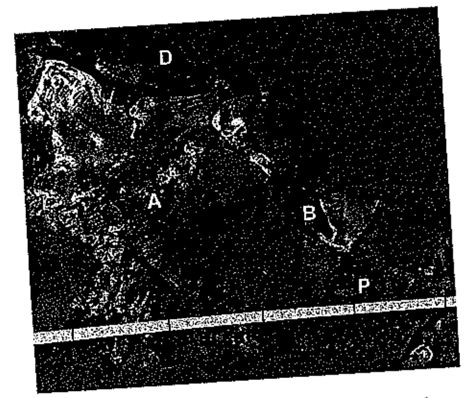

Plate 63 Large crystals of the glacier ice from bottom layers of Horn Glacier, Spitsbergen (intervals of 10 cm are marked). Note inclusions of air bubbles (A), cross section of the crystal basal planes (P), plastically deformed crystal boundaries (B) and presence of debris admixtures (D). Photo by Jacek Jania

convex shape and ice flow is predominantly vertical (downward) within it.

An ICE STREAM is a zone within an ice sheet or ice cap in which the ice flows significantly faster than on its sides. Theoretically, an ice stream has no rock boundaries; however, the course of the majority of ice streams is determined by their bedrock topography. Antarctic ice streams drain up to 90 per cent of mass accumulation in the interior (Paterson 1994: 301). The best known ice streams are on Siple Coast in Antarctica (Ice Streams A, B, C, D and E) and Jakobshavn Glacier, West Greenland. The measured velocities of Ice Stream B exceed 800 m a^{-1} whereas on the neighbouring ice sheet area they are less than 10 m a^{-1} (Van der Ween 1999). The surface velocity of Jakobshavn Glacier reaches almost 8 km per year near its terminus (Fahnestock *et al.* 1993: 1,532). In both cases, basal sliding is responsible for the fast flow of ice. High basal velocities of ice streams are thought to produce intense erosion of glacier beds.

In PERMAFROST areas several kinds of ground ice are present and ice bodies have a different origin, size, shape and internal structure. They might reasonably be classified into two genetic groups: (1) those originated from ground water and (2) those from water migrating from the surface to the ground. Within the first group, the largest ice masses are called hydro-laccolites. Their thickness can be as much as 20–40 m and their diameters can be dozens or even hundreds of metres (e.g. PINGO). Smaller forms are usually lenticular. *Ice lenses* appear in relatively dry ground or in sediments which have low permeability (e.g. tills). Owing to mineralization of water, which migrates within the sediments and to the higher than atmospheric pressure within the ground, crystallization of ice within the ground takes place in temperatures below the freezing point. Ice lenses grow upwards and have a horizontally laminated structure with inclusions of mineral particles. Ice wedges (see ICE WEDGE AND RELATED STRUCTURES) are the most common forms of the various ice bodies which result from water infiltration.

Ice jams on rivers most often take place in the winter and spring season and are a notable feature of large rivers which flow northwards in the northern hemisphere (Siberia, Canada). They are often caused by the earlier melting of snow in the southern parts of drainage basins and blockage of water by the longer preserved river ice cover in the northern reaches. Jamming of ice transport occurs as it lodges beneath the superficial ice cover of the frazil and anchor ice and thereby drainage blockage is typical for winter periods. The impediment may be so severe that wide areas of the adjacent floodplains become inundated. Detached drifting ice blocks may locally cause erosion of the floodplain sediments.

Specific ice masses originate in caves. The *ice cave* usually has a thick layer of laminated infiltration ice and icicles (ice stalactites and stalagmites). Ice caves are common in KARST areas within a permafrost zone. They are also known from locations where ground is not perennially frozen. In such cases, the very specific microclimate of the cave is responsible for the formation of annual infiltration ice layers: there is only one entrance located in the upper part of the cave and during winter, cold air flows down through the cave, while lighter warm air in spring and summer cannot penetrate into the system. Percolated meltwaters from snow or precipitation are frozen on the floor of the ice cave.

Tunnels formed in glaciers by englacial and subglacial drainage systems are called glacier caves. Vertical tunnels which transfer water from the glacier surface to subglacial channel are termed *ice shafts* or MOULINS. Ice shafts develop on planes of discontinuity within the glacier ice as crevasses and shear planes.

References

Dobrowolski, A.B. (1923) *Historja nauturalna lodu* [*Natural history of ice*], Warszawa: Kasa im. Mianowskiego.

Drewry, D. (1983) *Antarctica: Glaciological and Geophysical Folio*, Cambridge: Scott Polar Research Institute.

Fahnestock, M., Bindschadler, R., Kwok, R. and Jezek, K. (1993) Greenland Ice Sheet surface properties and ice dynamics from ERS-1 SAR imagery, *Science* 262, 1,530–1,534.

Jania, J. (1988) *Zrozumiec lodowce* [*Understanding glaciers*], Katowice: Wydawnictwo Slask.

Kotlyakov, V.M. (ed.) (1984) *Glyatsiologicheskiy slovar* [*Glaciological dictionary*], Leningrad: Gidrometeoizdat.

Menzies, J. (ed.) (2002) *Modern and Past Glacial Environments*, Oxford: Butterworth-Heinemann.

Paterson, W.S.B. (1994) *The Physics of Glaciers*, Oxford: Pergamon.

Shumskyi, P.A. (1955) *Osnovy structurnogo ledovedeniya* [*Principles of structural glaciology*], Moskva: Izdatelstvo Akademii Nauk SSSR; Trans. (1964) *Principles of Structrual Glaciology*, New York: Dover.

Titus, D.D., Larson, G.J., Strasser, J.C., Lawson, D.E., Evenson, E.B. and Alley, R.B. (1999) Isotopic composition of vent discharge from the Matanuska Glacier, Alaska: implication for the origin of basal ice, *Geological Society of America, Special Paper* 337, 37–44.

Van der Ween, C.J. (1999) *Fundamentals of Glacier Dynamics*, Rotterdam: A.A. Balkema.

Further reading

Hooke, R.L. (1998) *Principles of Glacier Mechanics*, Englewood Cliffs, NJ: Prentice Hall.

Martini, I.P., Brookfield, M.E. and Sadura, S. (2001) *Principles of Glacial Geomorphology and Geology*, Upper Saddle River, NJ: Prentice Hall.

Petrenko, V.F. and Whitworth, R.W. (1999) *Physics of Ice*, Oxford: Oxford University Press.

SEE ALSO: glacier; ice sheet; ice stream; ice wedge and related structures; permafrost

JACEK JANIA

ICE AGES (INTERGLACIALS, INTERSTADIALS AND STADIALS)

Ice ages have occurred throughout geological time: as, for example, during the Precambrian, the Ordovician, Permo-Triassic and the Cenozoic (Tertiary), of which the Quaternary is a part. During the Permo-Triassic ice age, the ice sheets were centred on the Antarctic continent, located across the southern polar area. Its principal glacial deposit, the Dwyka tillite, is found in South America, southern Africa, southern India and south Australia, a distribution that shows the fragmentation of the super continent (Gondwanaland) where glaciation occurred by 'continental drift'. The Cenozoic ice ages commenced when Antarctica experienced major glaciation about 34 Ma ago (Ma = millions of years). But, in a geomorphological sense, the term 'ice age', incorporating many ice ages and interglacials, most appropriately refers to the Quaternary ice ages, when major geomorphic changes occurred.

Evolution of the present boundary conditions of the Earth's climate, namely the location of the continents, seaways and mountain ranges, closely match the evolution of the climate system and ice ages. As the southern continents 'drifted' away from Gondwanaland, India collided with Asia to produce the largest landform on Earth, namely the Tibetan Plateau, a feature so high that it modulated the circulation of the high level westerlies that led to waves on the circum polar vortex. These drew cold arctic air to Canada and Europe. Cold continents adjacent to warm oceans were the ideal combination for initiating the major Quaternary ice ages. Once the closure of the Straits of Panama (about 4 Ma) had occurred, a zonal circulation of ocean water between the Atlantic and Pacific ceased, to be replaced by meridional warm water circulation in the Atlantic – the provider of precipitation to grow the ice sheets – and the start of the intensification of North Atlantic Deep Water formation, the driver of oceanic thermohaline circulation.

The standard definition of the Quaternary is that it commenced 1.8 Ma ago, a view tied to a stratotype (type-site) at Vrica, near Crotone, Calabria, Italy. This horizon lies at the top of the Olduvai magnetic reversal Subchron. But, an alternative view would place the start at about 2.5 Ma, because it coincides with the first record in marine sediments in the North Atlantic. This shows that ice sheets had grown large enough to reach tidewater lands where they launched icebergs into the ocean. When these melted, their load of ice rafted debris (IRD) was released onto the ocean floor. In China, it coincides with the contact between the 'Red Clay' and the first, Wucheng Loess; and also coincides with the base of the earliest, pre-Nebraskan, glaciation of North America.

The Quaternary ice ages are characterized by lowered temperatures when the snowline in mid-latitude regions was lowered by up to 1 km. This led to the growth of large mid-latitude ice sheets in Canada and Scandinavia, while the water they abstracted from the oceans lowered sea level by up to 135 m. But how did these large ice sheets grow, because they have no modern analogues? Three main theories have been proposed. (1) *Highland origin and windward growth*: based on the former Laurentide ice sheet in Canada and then applied to the Fenno-Scandian ice sheet. It is suggested that ice grew initially in the highland of Ungava and Quebec, then spread towards its source of precipitation from the Gulf of Mexico, before forming a large ice dome up to 3 km thick or more, centred over Hudson Bay. (2) *Instantaneous glacierization* involved rapid vertical growth of ice from large snow patches distributed over wide areas, including the low ground of Keewatin west of Hudson Bay. It led to a thinner multi-domed ice sheet, capable of rapid response to climate change. (3) *Marine-based ice sheets* are conceptually based on the West

Antarctic ice sheet and involved mid-latitude ice sheets centred on shallow sea areas such as Hudson Bay or the Baltic Sea, where large ice domes grew. These, drained by ice streams, were buttressed by floating ice shelves which, on destabilization, no longer provided support, so that the ice domes collapsed and surged into the ocean. Whatever the mode of ice sheet growth, it is clear that they were large enough to reach the tidewater regions of the North Atlantic Ocean on several occasions when they launched armadas of icebergs. On melting, these released the glacial material they were carrying which was deposited in the ocean as discrete layers of ice rafted debris (IRD). Such episodes are known as Heinrich Events and they occurred on an irregular 5 to 7 ka cycle.

South of the ice sheets vegetation zones were compressed towards the equator. Accompanying this was a migration of the fauna and flora and in geomorphic processes. Mid-latitudes experienced periglacial climates and atmospheric circulation was more vigorous as the consequence of steeper poleward climate gradients. Low latitudes experienced widespread aridity. Ice ages were diversified by brief climatic ameliorations called interstadials separated by colder stadials. Interglacials supported no mid-latitude ice sheets, had relatively high sea levels and a fauna and flora similar to the present (Holocene) interglacial. During the past million years or so, ice ages have commonly lasted about 100,000 years. But the length of interglacials appears to have been variable and is controversial. The interglacial about 400,000 years ago may have lasted for the best part of 60,000 years; but duration of the 'last interglacial' about 125 ka is controversial: according to some, it lasted about 10,000 years, but others believe it was twice as long. The current interglacial has already lasted 11,600 years, during which time several minor fluctuations in climate have occurred, such as the Little Ice Age.

Figure 91 shows that some fifty ice ages and fifty interglacials occurred in the last 2.5 Ma. This contrasts with earlier, classical, views that maintained that only four ice ages occurred in the 'Great Quaternary Ice Age', a view based on the record of Alpine glacial advances shown by four outwash terraces in Bavaria by Penck and Bruckner in 1910. Their view was reinforced in America and elsewhere by four major groups of glacial deposits.

Their paradigm lasted for over sixty years. But evidence of greater complexity came from an unexpected source, namely marine deposits of the deep open ocean, where long sequences recorded the ice ages of the entire Quaternary. These consist of muds composed of microfossils of plankton (planktic) and benthos (benthic) which secreted their calcareous shells in oxygen isotopic equilibrium with the ocean water they inhabited. Initially it was believed that variability in the ratio of ^{18}O and ^{16}O in foraminifera microfossils was primarily an indicator of the ocean temperatures in which the organisms grew, and only to a lesser extent was the isotopic composition of the ocean involved. Subsequently it was shown that the isotopic composition, not the temperature, of the ocean was the primary control on oxygen isotope

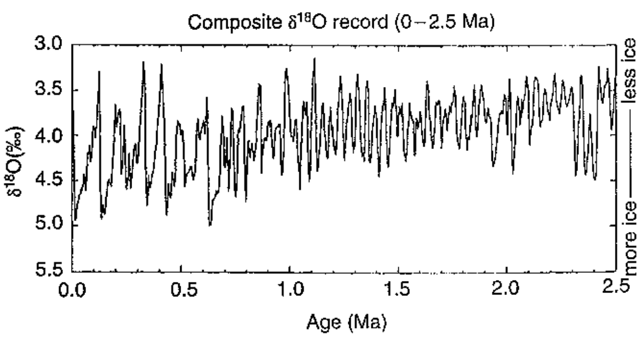

Figure 91 Composite record of the ice ages for the last 2.5 million years, based on oxygen isotope variability (from Crowley and North 1991)

ratio. Furthermore, because bottom water temperatures remained constantly cold, then benthic foraminifera would only record the oxygen isotopic composition of the ocean. Thus, because the primary control on this was the volume of isotopically light ice locked up on the continents during ice ages, then the benthic signal would provide an ice volume indication: that is, a record of the ice ages and interglacials. Moreover, because this was altered as the sea rose and fell as continental ice volume changed, the benthic signal was also one of sea-level change. Recently, it was shown that bottom water temperatures had varied; thus ^{18}O benthic ratios could no longer provide precise ice volume or sea-level information. Fortunately this was resolved by temperature determinations based on Magnesium/Calcium ratios in Ostracoda, that allowed discrete ice volume and temperature parameters to be identified separately.

Oxygen isotope signals are subdivided into *oxygen isotope stages* and *sub-stages*. Interglacials are numbered backwards in time with odd numbers, starting with the present (Holocene) interglacial as stage 1. Similarly ice ages are numbered backwards with time with even numbers, starting with the last ice age or Last Glacial Maximum as stage 2. Considerable confusion has been caused by the designation of an oxygen isotope stage 3 in the middle of the last 100 ka cycle. This arose because it was once thought that the 41 ka tilt frequency was the main ice age pacing (below). So in reality, the present (Holocene) interglacial is stage 1 and the last interglacial centred on ~125 ka is sub-stage 5e, the warmest part of stage 5, when conditions more or less corresponded with the present. Stage 3 is not an interglacial but merely a general climatic amelioration punctuated by millennial scale interstadials (below) during the last glacial cycle.

A means of providing a chronology for the ice ages and interglacials was provided by fixed points in the cores where changes in the Earth's magnetism occurred. Of these, the most important reversal is the Matuyama/Brunhes reversal at 0.78 Ma (780 ka). By assuming constant sedimentation, ages were provided for the ice age and interglacial boundaries. This was reinforced when it was found that the predicted age of the 'last interglacial' of 125 ka was supported by Uranium 234–Thorium 230 ages on uplifted coral reefs that showed a relatively high sea level at that time. Since then, the same has been confirmed for the age of the interglacial at about 300 ka, oxygen isotope stage 9. The last interglacial shoreline is found throughout the world as coral reefs, raised beaches or shore-platforms. Detailed records of sea-level fluctuations have been established in regions with uplifted coral reefs, such as Barbados, Tahiti and the Huon Peninsula of New Guinea.

Inspection of oxygen isotope signal reveals a number of cycles or pacings (Figure 91). These are the well-known pacings caused by the orbital movement of the Earth around the sun, and occur at 100 ka (eccentricity of the orbit), 41 ka (variation of the ecliptic or tilt of the Earth's axis), 23 ka and 19 ka (precession of the equinoxes that varies the distance between the Earth and the sun at the summer solstice). Relating the changing amount of solar energy received at given latitudes to the oxygen isotope (ice volume) record, however, had to await confirmation of these pacings in high sedimentation marine cores from the Indian Ocean. The astronomical and oxygen isotope records are not an exact match, because of a time-lag between orbital forcing and the climatic response as ice sheets grew or decayed to change the oxygen isotopic composition of the ocean. By using phase relationships (leads and lags), the oxygen isotope record has been 'tuned' by the predicted orbital changes. This controversial method is illustrated by the debate over the commencement age and duration of the 'last interglacial' (oxygen isotope sub-stage 5e) which Uranium–Thorium ages on stalagmite shows has been underestimated. It may be resolved because the shorter estimate is based on orbital calculations for 65°N (the 'sensitive latitude' for mid-latitude ice sheets), whereas the longer one is more consistent with orbital calculations for 65°S.

Despite ongoing debate about the precision of ice age and interglacial timing, it would seem that the origin of the ice ages has been discovered. Changes in temperature caused by changes in the orbit of the Earth seem to be the pacemaker of climate change. But how was orbital forcing transformed into actual climate change? Several questions remain before the matter can be settled. The earlier part of the Quaternary record displays a strong 41 ka tilt pacing with no trace of the 100 ka eccentricity pacing that only becomes distinct after 700 ka (Figure 91). This was when the large mid-latitude ice sheets grew and extended south to Ohio, London, Berlin and Kiev for the first time. What was the cause of the later

Quaternary 100 ka cycles? Perhaps further uplift of the Tibetan Plateau intensified the climate forcing? More likely is that continental erosion had progressively removed extensive layers of weathered rock to reveal more resistant bedrock. Low gradient ice sheets occur on soft deformable beds, but more resistant rocks support ice domes with steep marginal profiles. These depressed the Earth's crust, while continuing to receive more snow at their summits, thus prolonging the ice age. This may account for the transition from the 41 ka to the 100 ka world. Not unrelated to this is the large mismatch between the forcing provided by the 100 ka pacing and the response of the climate system (Figure 92). While there is a predictable and strong response of the climate system in the tilt and precessional bands, the disproportionate response in the eccentricity band remains unexplained. Could it be because of the development of large ice domes (above), or perhaps the real role of eccentricity is to modulate the tilt and precession forcings?

Amplification of orbitally forced climate change occurred by enhanced or decreased greenhouse gas concentrations of carbon dioxide and methane in the atmosphere. These are clearly implicated in climate change as is shown by measurements of past atmospheres preserved within bubbles of air trapped in ice layers of the Antarctic and Greenland ice sheets (Figure 93). Higher concentrations of greenhouse gases correspond with interglacials, while they are lowered during the ice ages. Oxygen isotope or Deuterium analyses of ice from cores through the Greenland and Antarctic (Figure 93j) ice sheets, provide a record of temperature changes at the surface of the ice sheet. But the age of the ice layers and the bubbles of air within them is not the same because air bubbles provide a greenhouse gas record at the time when they were sealed by the weight of overlying snow and firn. Differences of up to sixty years occur in high snowfall regions such as Greenland, but up to 1,200 years in the centre of Antarctica where only about 2 cm of snow falls

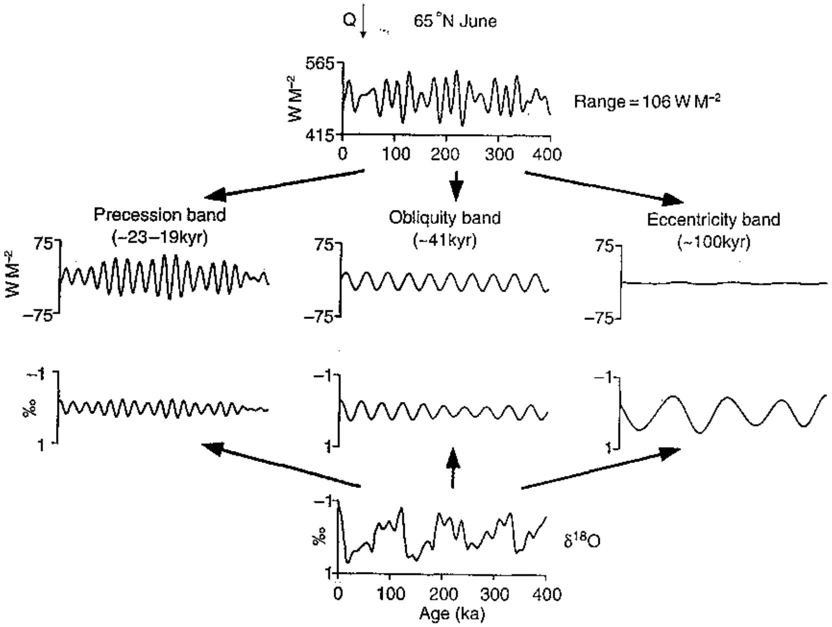

Figure 92 The 100,000 cycle problem shown by partitioning the radiation and climate time series into their dominant periodic components (from Imbrie *et al.* 1993)

annually. This may conceal important phase relationships between temperature and greenhouse gases. Fortunately the sharp methane spikes in the record allow good correlation between interhemispheric ice cores. The record of temperature and greenhouse gases now extends back more than 400,000 years at Vostok Station, Antarctica, through four major ice ages and five interglacials; whereas the record at Greenland Stations only extends back beyond the last interglacial.

Not only do palaeotemperature records from ice sheets reveal the main orbital pacings, they also display strong millennial ones (Figure 93). Moreover, unlike the sinusoidal shape of the

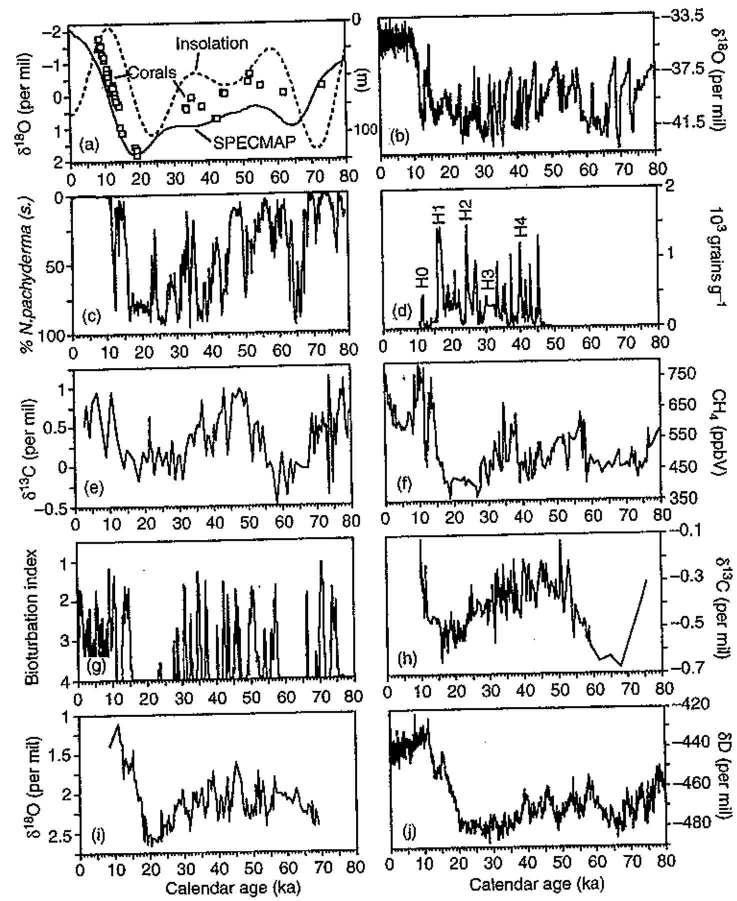

Figure 93 Millennial pacings (Alley and Clark 1999). (a) marine oxygen isotope signal (SPECMAP), insolation (orbital forcing) and sea level from dated corals; (b) Greenland ice sheet project (GRIP) temperatures; (c) sea surface temperatures from *N. pachyderma*, off western Ireland; (d) ice rafted debris and Heinrich Events; (e) tropical Atlantic 13C variability; (f) methane concentrations in the Greenland Ice Sheet Project 2 (GISP2); (g) changes in sediment type, Santa Barbara Basin, California; (h) north-east Pacific 13C variability; (i) southern ocean sea surface temperatures from oxygen isotopes; (j) deuterium (temperature) variability, Vostok Station, South Pole

orbital pacings, the millennial ones are square-waved in shape, and show abrupt changes in temperature. Some of these represent changes in temperature of up to 10 °C in less than a decade. Matching millennial records from high sedimentation marine core records also correspond with sea surface temperature variability (Figure 93c). These are estimated from the palaeoecological requirements of planktic fossils, notably the polar *Neogloboquadrina pachyderma* (with sinistral coiling). These millennial records are joined by other records: for example, continental pollen records in Europe, by the monsoon record from the Arabian Sea, sea surface temperatures from the Sulu and China Seas, changes in biological productivity or ocean ventilation inferred from ^{13}C millennial variability (Figure 93h), and changes in water masses from offshore California (Figure 93g).

Such millennial pacings are global in character and they show that the Earth's climate changed on short and abrupt timescales. The principal pacing, that underpins all the major changes in climate over the past 110,000 years, is that at 1,450 (sometimes stated as 1,500) years. It corresponds to pacings in the production rate of the cosmogenic isotope Berylium 10 (^{10}Be) also recorded in ice sheets. ^{10}Be is produced in the atmosphere by cosmic ray bombardment. Higher quantities are produced when the solar magnetic field is at its weakest and not shielding the Earth's atmosphere from cosmic ray bombardment. Conversely, lower quantities are produced when the solar magnetic field is at its strongest and shielding the Earth from cosmic bombardment. Therefore, lower quantities of ^{10}Be correspond to a stronger and warmer Sun and vice versa. Thus variability in solar output appears to control the 1,450 climate cycle. The 1,450 pacing occurs throughout the last glacial cycle as well as the current (Holocene) interglacial. Its amplitude is greater during the unstable geographical configuration of the ice ages because of large ice sheets, low sea levels and an isostatically depressed crust. But its interglacial mode is manifested in cycles of sea ice formation in the North Atlantic, the record of alpine glacier fluctuation, and in high sedimentation continental records throughout the world.

Cycles of ice core interstadials are known as Daansgard–Oeschger cycles, named after the Danish and Swiss scientists who discovered them. Some of them correspond to the less perfect record of interstadials on the continents, where evidence from pollen and insect faunas show climatic amelioration during ice ages. The Bølling–Allerød interstadial, between about 15 and 13 ka ago, is well known and was followed by the Younger Dryas stadsial. This dramatic reversal to cold conditions just before the Holocene may have been forced by a major meltwater flux to the North Atlantic from glacial Lake Aggasiz in North America, or it may be the result of solar forcing. Cycles of Heinrich Events (Figure 93) are known as Bond cycles, and they contain packages of Dansgaard–Oeschger interstadials. Heinrich Events may have been caused by mechanisms internal to the ice sheets, but their synchronous nature around the North Atlantic suggests they may have been forced by changes in climate.

Comparison of Greenland and Antarctic ice cores shows broad similarities, but also some important differences. One school of thought believes that deep water thermohaline ocean circulation in the Atlantic carries climate signals between the two hemispheres and is a prime control on the global climate system, with the signals originating in the North Atlantic. Others see the climate signal originating in the tropical Pacific Ocean, its global transmission taking place as millennial scale phenomena similar to El Niño events.

D.Q. BOWEN

References

Alley, R.B. and Clark, P.U. (1999) The Deglaciation of the Northern Hemisphere: A Global Perspective, *Annual Reviews of Earth and Planetary Sciences* 27, 149–182.

Crowley, T.J. and North, G.R. (1991) *Paleoclimatology*, New York and Oxford: Oxford University Press.

Imbrie, J. Berger, A., Boyle, E.A., Clemens, S.C., Duffy, A., Howard, W.R. *et al.* (1993) On the structure and origin of major glaciation cycles, 2: the 100,000 cycle, *Paleoceanography* 8, 699–736, American Geophysical Union.

Further reading

Alverson, K.D., Bradley, R.S. and Pederson, T. (eds) (2003) *Paleoclimate, Global Change and the Future*, Berlin: Springer.

Bradley, R.S. (1999) *Paleoclimatology*, London: Academic Press.

Clark, P.U., Webb, R.S. and Keigwin, L.D. (1999) *Mechanisms of Global Change at Millennial Time Scales*, American Geophysical Union Monograph 112, Washington, DC.

Cronin, Thomas M. (1999) *Principles of Paleoclimatology: Perspectives in Paleobiology and Earth History*, New York: Columbia University Press.

ICE DAM, GLACIER DAM

Glacier dams are barriers of ice that act as a seal for water impounding glacial lakes. Water may be impounded within the GLACIER (*englacially* and *subglacially*), on the surface of the glacier (*supraglacially*), or in lakes formed at the glacier's edge (*marginally* or *proglacially*). Subaerial glacial lakes (marginal or proglacial lakes) are confined by ice on one side and by topographical barriers on the other. Subglacial lakes may form a cupola above the glacier bed at subglacial geothermal areas, accompanied by a depression in the glacier's surface, or be situated in a bedrock hollow beneath a relatively flat dome-shaped glacial surface. Supraglacial lakes are isolated in depressions on the glacier surface.

In general, ice-dammed lakes of every type can drain in episodic bursts. The glacier's surface and the water pressure potential slope towards the lake. The ice-dammed lakes receive water inflow and are gradually made to expand. Basal water pressure will increase and the lake level will be raised. Eventually, the hydraulic seal of the ice dam will be ruptured at the glacier base, the hydraulic seal opened, and seepage causes enlargement of the drainage system, initiating a flood under the surrounding ice. After discharge has begun, pressure from the ice constricts the passageway, and water flow at an early stage in the flood correlates primarily with enlargement of the ice tunnel due to heat from friction against the flowing water and to thermal energy stored in the lake. Increasing as an approximate exponent of time over a matter of hours or days, the discharge falls quickly after peaking. The recession stage of the hydrograph sets in when tunnel deformation begins to exceed enlargement by melting. Fluctuations in the thickness of the blocking ice, due to climatic variations or surges, may modify the outburst cycle or even stop bursts completely.

Occasionally, glacial outbursts are triggered by flotation of the ice dam. Rather than initial drainage from the lake being localized in one narrow conduit, the water is suddenly released as a sheet flow, surging downhill and propagating a subglacial pressure wave, which exceeds the ice overburden and lifts the glacier in order to create space for the water. In this instance, discharge increases faster than can be explained by conduits expanding due to melting.

Marginal lakes at subpolar glaciers, where the ice barrier is frozen to the bed, are typically breached as water spills over the top of the dam into supraglacial channel that melts into a bigger breach – commonly at the juxtaposition of the glacier and a rock wall.

Glacial outbursts (*jökulhlaup* in Icelandic) can have pronounced geomorphological impact, since they scour river courses and inundate floodplains. Outbursts result in enormous erosion, for they carry huge loads of sediment and imprint the landscape, with deep canyons, channelled SCABLANDS, ridges standing parallel to the direction of flow, sediment deposited on outwash plains, coarse boulders strewn along riverbanks, kettleholes where massive ice blocks have become stranded and melted, and breached terminal moraines. Some modern outbursts have produced flood waves in coastal waters (TSUNAMIS). In the North Atlantic, outburst sediment dumped onto the continental shelf and slope has been transported far away by turbid currents. Outburst floods wreak havoc along their paths, threatening people and livestock, destroying vegetated lowlands, devastating farms, disrupting infrastructure such as roads, bridges and power lines, and threatening hydroelectric plants on glacially fed rivers.

Further reading

Tweed, F.S. and Russell, A.J. (1999) Controls on the formation and sudden drainage of glacier-impounded lakes: implications for jökulhlaup characteristics, *Progress in Physical Geography* 23, 79–110.

HELGI BJORNSSON

ICE SHEET

An ice sheet is a large dome-shaped ice mass (> 50,000 km²) in area) that exhibits a generally radial flow pattern which is predominantly unconstrained by the underlying topography. If the ice mass is less than 50,000 km², it is usually referred to as an ice cap. Ice sheets are comprised of a single ice dome or series of coalescent domes that represent the highest parts of the ice sheet surface. In the interior regions of the three present-day ice sheets in East Antarctica, West Antarctica and Greenland, ice thickness often exceeds several thousand metres. These central accumulation areas (or dispersal centres) are marked by ice divides which delineate neighbouring catchment areas (analogous to a drainage divide in fluvial systems).

Towards the margins of an ice sheet, the underlying topography becomes increasingly more important in channelling flow and fast moving outlet GLACIERS and ICE STREAMS may develop in deep subglacial troughs. Other ice streams may arise irrespective of the bedrock topography. Thus, flow within an ice sheet is generally divided into slow 'sheet' flow, typical of ice domes, and fast 'stream' flow, which occurs in outlet glaciers and ice streams. This may be an oversimplification because some ice streams and outlet glaciers have smaller tributaries of intermediate velocity extending well into the ice sheet interior. Ice streams and outlet glaciers are known to be a key control on overall ice sheet stability because of their capacity to rapidly drain large volumes of ice.

Once established, ice sheets respond to climate forcing and also influence the climate system over a global as well as local scale. They represent a major obstacle to atmospheric circulation and produce some of the largest regional anomalies in albedo and radiation balance (Clark *et al.* 1999). Ice sheets also store and release considerable amounts of freshwater and are one of the main regulators of continental water balance and global sea-level change. For example, during the Last Glacial Maximum (18,000–21,000 yr BP), global SEA LEVEL was around 120 m lower as continental ice sheets developed in the high to mid-latitudes of the northern hemisphere, covering large parts of North America (e.g. Laurentide Ice Sheet) and Europe (e.g. British and Scandinavian Ice Sheets). It has been discovered that rapid discharges of ICEBERGS and meltwater (see MELTWATER AND MELTWATER CHANNEL) associated with these former ice sheets exerted a profound and often abrupt impact on ocean circulation and climate (cf. Clark *et al.* 1999). There is also compelling evidence that some former ice sheets (e.g. Laurentide Ice Sheet) were characterized by relative instability during the last glacial cycle, indicating that even the largest ice masses are highly dynamic systems (Boulton and Clark 1990). A major challenge for contemporary ice sheet research lies in predicting the response of the potentially susceptible West Antarctic Ice Sheet to future changes in sea level and climate, particularly global warming.

References

Boulton, G.S. and Clark, C.D. (1990) A highly mobile Laurentide ice sheet revealed by satellite images of glacial lineations, *Nature* 346, 813–817.

Clark, P.U., Alley, R.B. and Pollard, D. (1999) Northern Hemisphere ice-sheet influences on global climate change, *Science* 286, 1,104–1,111.

Further reading

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

Paterson, W.S.B. (1994) *The Physics of Glaciers*, 3rd edition, Oxford: Elsevier Science.

SEE ALSO: glacier; glacial deposition; glacial erosion; glacial theory; ice

CHRIS R. STOKES

ICE STAGNATION TOPOGRAPHY

Often also referred to as 'dead-ice topography', this is undulatory to hummocky terrain composed of a wide range of glacigenic sediments that accumulate on the surface of a melting glacier by the processes of melt-out, mass movement and glacifluvial reworking (Boulton 1972). Hummocks are often randomly distributed and of different dimensions, giving a chaotic appearance, but some linearity may be imparted by former ice structures. Although the term ice stagnation topography has been traditionally associated with deglaciated terrain, large areas of such supraglacial landform assemblages often still contain buried glacier ice, sometimes many thousands of years in age. In glaciers that carry relatively large debris loads, englacial and supraglacial sediment-landform associations develop over long periods of time and are subject to numerous cycles of reworking, re-mobilization and topographic inversion before they are finally lowered to the glacier substrate. The wide range of processes involved in the production of ice stagnation topography, including mass flowage, meltwater reworking, lacustrine sedimentation folding and faulting, result in the occurrence of a variety of sediment assemblages.

Although the term 'hummocky moraine' has been used to describe a wide range of landforms, it has most recently been restricted to moundy, irregular topography deposited by the melt-out of debris-mantled glaciers and therefore also relates to stagnating ice but is not strictly synonymous with the term ice stagnation topography. For example ice stagnation topography often encompasses large areas of glacifluvially reworked materials or supraglacial KAME and kettle (see KETTLE AND KETTLE HOLE) topography and even complex ESKER systems and ice-walled lake plains. In some circumstances the melt-out process results in the partial preservation of englacial structures, specifically alternating debris-rich and debris-poor ice layers, and the production of transverse linear elements called 'controlled moraine'. The preservation potential of controlled moraines is poor but they are very striking elements of freshly deglaciated terrain at high latitudes where glacier snouts are melting down very slowly and supraglacial sediment reworking is locally restricted.

The processes of differential ablation, multicyclic debris reworking and topographic inversion and complex meltwater drainage development on a debris-rich glacier will result in the production of a thick supraglacial debris mantle that can decouple glacier response to climate forcing. Ice wastage or stagnation can therefore lag behind climatic inputs by decades or even thousands of years. As meltwater systems gradually open up large subglacial and englacial drainage conduits so the developing ice stagnation topography may become perforated by numerous moulins, ponds and lakes. Referred to as glacier karst (Clayton 1964), these water-filled depressions often coalesce to produce large supraglacial lakes.

References

Boulton, G.S. (1972) Modern arctic glaciers as depositional models for former ice sheets, *Journal of the Geological Society of London* 128, 361–393.

Clayton, L. (1964) Karst topography on stagnant glaciers, *Journal of Glaciology* 5, 107–112.

Further reading

Benn, D.I. (1992) The genesis and significance of 'hummocky moraine': evidence from the Isle of Skye, Scotland, *Quaternary Science Reviews* 11, 781–799.

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

Eyles, N. (1979) Facies of supraglacial sedimentation on Icelandic and alpine temperate glaciers, *Canadian Journal of Earth Sciences* 16, 1,341–1,361.

Fleisher, P.J. (1986) Dead-ice sinks and moats: environments of stagnant ice deposition, *Geology* 14, 39–42.

Gravenor, C.P. and Kupsch, W.O. (1959) Ice disintegration features in western Canada, *Journal of Geology* 67, 48–64.

Gustavson, T.C. and Boothroyd, J.C. (1987) A depositional model for outwash, sediment sources, and hydrologic characteristics, Malaspina Glacier, Alaska: a modern analog of the southeastern margin of the Laurentide Ice Sheet, *Geological Society of America Bulletin* 99, 187–200.

Ham, N.R. and Attig, J.W. (1996) Ice wastage and landscape evolution along the southern margin of the Laurentide Ice Sheet, north-central Wisconsin, *Boreas* 25, 171–186.

Huddart, D. (1999) Supraglacial trough fills, southern Scotland: origins and implications for deglacial processes, *Glacial Geology and Geomorphology* http://boris.qub.ac.uk/ggg/papers/full/1999/rp04.html

Johnson, M.D., Mickelson, D.M., Clayton, L. and Attig, J.W. (1995) Composition and genesis of glacial hummocks, western Wisconsin, USA, *Boreas* 24, 97–116.

Nakawo, M., Raymond, C.F. and Fountain, A. (2000) *Debris-Covered Glaciers*, IAHS Publication no. 264.

Paul, M.A. (1983) The supraglacial landsystem, in N. Eyles (ed.) *Glacial Geology*, 71–90, Oxford: Pergamon.

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

Thomas, G.S.P., Connaughton, M. and Dackombe, R.V. (1985) Facies variation in a late Pleistocene supraglacial outwash sandur from the Isle of Man, *Geological Journal* 20, 193–213.

DAVID J.A. EVANS

ICE STREAM

An ice stream is a zone of an ICE SHEET that flows much faster than the surrounding ice. Typical ice sheet velocities are of the order of tens of $\mathrm{m\,a^{-1}}$ but ice stream velocities range from hundreds to several thousands of $\mathrm{m\,a^{-1}}$. Ice streams display a broad range of characteristics and behaviour but they are generally large features with widths of tens of km and lengths of hundreds of km. Their large size and profligate ice flux dominate ice sheet discharge and it is for this reason that they are viewed as a critical control on ice sheet mass balance and stability.

A defining characteristic of ice streams is that they are bordered by slower moving ice. This creates heavily crevassed lateral shear margins that aid their identification. If the fast-flowing ice is bordered at the surface by rock walls, it is usually referred to as an outlet glacier. It should be noted, however, that many ice streams show characteristics of both along their length. To add to their complexity, some ice streams appear to be fed by numerous smaller tributaries that penetrate up to 1,000 km into the interior of the ice sheet (Bamber *et al.* 2000).

Ice streams have been investigated in the East Antarctic, West Antarctic and Greenland Ice Sheets and have also been hypothesized in many palaeo-ice sheets. They occur in a variety of settings and their fast flow may be achieved through a variety of flow mechanisms. At one end of the spectrum, some ice streams occupy deep bedrock troughs and they are often characterized by steep surface slopes and high driving stresses. A large component of their fast flow arises from rapid basal sliding but, under exceptionally high driving stresses, thermally enhanced deformation of a thick basal ice layer may be the dominant flow mechanism (cf. Iken *et al.* 1993). In contrast, some ice streams have low surface slopes, low driving stresses and do not appear to be constrained by the underlying topography. Their behaviour is more enigmatic but their fast flow appears to be related to a metres thick layer of soft, saturated sediment that deforms beneath the ice stream and/or provides an effective surface for basal sliding (Engelhardt and Kamb 1998). Other ice streams show characteristics of both of these end-member behaviours and many more ice streams remain to be studied.

References

Bamber, J.L., Vaughan, D.G. and Joughin, I. (2000) Widespread complex flow in the interior of the Antarctic Ice Sheet, *Science* 287, 1,248–1,250.

Engelhardt, H. and Kamb, B. (1998) Basal sliding of Ice Stream B, West Antarctica, *Journal of Glaciology* 44, 223–230.

Iken, A., Echelmeyer, K., Harrison, W. and Funk, M. (1993) Mechanisms of fast flow in Jakobshavns Isbræ, West Greenland: Part I. Measurements of temperature and water level in deep boreholes, *Journal of Glaciology* 39, 15–25.

Further reading

Bentley, C.R. (1987) Antarctic ice streams: a review, *Journal of Geophysical Research* 92, 8,843–8,858.

Stokes, C.R. and Clark, C.D. (2001) Palaeo-ice streams, *Quaternary Science Reviews* 20, 1,437–1,457.

SEE ALSO: glacier; ice; ice sheet

CHRIS R. STOKES

ICE WEDGE AND RELATED STRUCTURES

Ice wedges are among the most common forms of ground ice within continuous PERMAFROST. They are commonly 1.0–1.5 m wide at the top and up to 4 m deep (Plate 64), although extreme examples up to 4 m wide and 10 m deep are reported from Siberia. In plan the linear ice wedge structures link to form polygonal patterns with a mesh size of several tens of metres, termed ice-wedge polygons or tundra polygons. The surface expression is a linear furrow due to differential settlement of the active layer immediately above the ice wedge.

Ice wedges result from thermal contraction cracking of the permafrost. Low winter air temperatures, and lack of insulating snow, lead to rapid cooling of the ground and the development of tensile stresses. Lachenbruch (1962) showed that cracking is most likely when the ground temperature is below approximately −20 °C, though slightly lower temperatures may be necessary in sands and gravels than is the case in silts and clays. Entry of snow and hoarfrost prevents crack closure as the permafrost warms during the following summer, leaving thin (often less than 1 mm; Mackay 1992) veins of ice penetrating into the permafrost. These provide lines of weakness for further cracking in subsequent years, leading to progressive widening into ice-wedges over time. Wedge growth associated with individual cracking events is generally less than 1 mm, and cracking may occur only periodically, so that a 1-m wide ice wedge may take thousands of years to form (Harry and Gozdzik 1988).

Epigenetic ice wedges develop below a stable ground surface and are younger than the adjacent frozen host sediments, while syngenetic ice wedges form contemporaneously with the slow accumulation of host sediments in subaerial permafrost environments. Where aeolian sediment transport is active, sand or silt – rather than snow and hoarfrost – may enter open thermal contraction cracks, resulting in the formation of sand-filled wedge structures termed sand wedges (Murton *et al.* 2000).

Plate 64 Upper portion of an ice wedge, Ellesmere Island, Canada. Note the active layer above the wedge is approximately 0.75 m deep at this location

A mean annual air temperature value of between −6 °C and −8 °C has been widely used as the warm limit for the development of thermal contraction cracking and the formation of ice wedges (Péwé 1966), although variations in seasonal extreme temperatures may lead to a range of limiting mean annual temperatures. Thawing of permafrost and contained ice wedges is associated with slumping of sediments to fill the resulting voids, thus preserving the wedge forms within the host sediments. Such 'pseudomorphs' or 'casts' are important stratigraphic markers in sedimentary sequences, providing evidence for the former existence of permafrost (Svensson 1988). Host sediments may be upturned against the sides of casts, marking former compression during summer ground warming, or there may be downward slumping of adjacent sediments into the cast.

Ice-wedge casts and sand wedges may be preserved within sedimentary units (intraformational), between sedimentary units (interformational) or they may penetrate downwards from the present ground surface (supraformational). Intraformational casts suggest episodic sediment accumulation in a permafrost environment and are common in fluvial gravel trains (Seddon and Holyoak 1985). Interformational wedge casts indicate a major change in sedimentary environment separated by a phase of permafrost. Supraformational casts may be visible in aerial photographs as polygonal 'crop marks' (Svensson 1988).

References

Harry, D.G. and Gozdzik, J.S. (1988) Ice wedges: growth, thaw transformation and palaeoenvironmental significance, *Journal of Quaternary Science* 3, 39–55.

Lachenbruch, A.H. (1962) Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost, *Geological Society of America Special Paper* 70.

Murton, J.B., Worsley, P. and Gozdzik, J. (2000) Sand veins and wedges in cold Aeolian environments, *Quaternary Science Reviews* 19, 899–922.

Mackay, J.R. (1992) The frequency of ice-wedge cracking (1967–1987) at Garry Island, western Arctic coast, *Canadian Journal of Earth Sciences* 29, 236–248.

Péwé, T.L. (1966) Palaeoclimatic significance of fossil ice-wedges, *Biuletyn Peryglacjalny* 15, 65–73.

Seddon, M.B. and Holyoak, D.T. (1985) Evi of sustained regional permafrost during depo of fossiliferous Late Pleistocene sediments at Stanton Harcourt, Oxfordshire, England, *Proceedings of the Geologists' Association* 96, 53–71.

Svensson, H. (1988) Ice-wedge casts and relict polygonal patterns in Scandinavia, *Journal of Quaternary Science* 3, 57–67.

Further reading

French, H.M. (1996) *The Periglacial Environment*, 2nd edition, Harlow: Longman.

CHARLES HARRIS

ICEBERG

Icebergs are floating pieces of glacier ice that are found in marine and lake waters. They are produced when ice breaks off from the margins of glaciers and ice sheets that terminate as ice cliffs in the sea and lakes. Icebergs vary in length from metres to kilometres – those less than 5–10 m in length are referred to as bergy bits. Larger icebergs are usually flat-topped or tabular in shape, whereas smaller bergs, and those in the later stages of melting and breakup, are of irregular form. The ice in icebergs comes originally from snowfall on the glacier surface and is transformed into glacier ice as a result of pressure during burial. This ice may be hundreds or even thousands of years old by the time it calves into the sea. Icebergs are less dense than water because of the presence of air bubbles that are trapped within them during formation. This density difference makes them buoyant, and about 80–90 per cent of their bulk is hidden below the water surface. Iceberg keels can reach hundreds of metres deep. Once icebergs have broken off from their parent glacier, they drift under the influence of ocean currents and, to a lesser extent, wind.

Reference

Wadhams, P. (2000) *Ice in the Ocean*, Amsterdam: Gordon and Breach.

JULIAN A. DOWDESWELL

ICING

Sheets of ice formed on the ground or on lake or river ice by the freezing of successive flows of water discharging from the ground or rising up through fractures in an antecedent ice cover. Icings are not restricted to PERMAFROST areas, but are common in the discontinuous permafrost zone, especially in carbonate terrain. Icings may be a serious hazard for road traffic, occurring persistently at the base of road cuts and at creek crossings. Most icings melt during the first few weeks of summer, but some extensive features may persist for years.

River icings are the most extensive features, and may extend over tens of km^2. The spatial extent of the icing depends on the discharge rate of water to the ice surface, water temperature, air temperature and channel slope (Hu and Pollard 1997). Icings are characteristically layered, with each bed the product of a discharge event. The basal portion of these layers may be discoloured by the solutes expelled and concentrated during freezing. Icings may extend outside the channel used by the river in summer, and may leave a dusting of precipitate on the local vegetation as they melt.

Icing blisters form if hydraulic pressure lifts an overlying ice layer. In continuous permafrost, such blisters have been observed at the edges of pingos and in residual ponds of drained lake beds, where water is expelled during permafrost aggradation (Mackay 1997). Such features are similar to frost blisters, but lack overlying ground.

References

Hu, X. and Pollard, W.H. (1997) The hydrologic analysis and modelling of river icing growth, North Fork Pass, Yukon Territory, Canada, *Permafrost and Periglacial Processes* 8, 279–294.

Mackay, J.R. (1997) A full-scale experiment (1978–1995) on the growth of permafrost by means of lake drainage, western Arctic coast: a discussion of the method and some results, *Canadian Journal of Earth Sciences* 34, 17–33.

C.R. BURN

ILLUVIATION

Illuviation is a process by which material removed from one horizon is deposited in another horizon of a soil (Foth 1984; Ritter 1986; Soil Science Society of America 1987). Usually the direction of transport is from an upper to a lower horizon. The lower horizon, an illuvial horizon, is considered a zone of accumulation and concentration for material. The material can be precipitated from solution or deposited from suspension. Depending on the conditions of soil formation, different illuviated constituents are present. For example, an acid soil formed under forest vegetation may exhibit a dark-coloured illuvial horizon containing quantities of sesquioxides and mineral-organic complexes as well as clay minerals and clay-size material. These materials have been eluviated or transported from an overlying horizon that may be significantly depleted in these constituents. A soil in a prairie region may exhibit only clay increases in the illuvial horizon or zone of accumulation. For contrast, see ELUVIUM AND ELUVIATION.

References

Foth, H.D. (1984) *Fundamentals of Soil Science*, New York: Wiley.

Ritter, D.F. (1986) *Process Geomorphology*, 2nd edition, Dubuque, IA: William Brown.

Soil Science Society of America (1987) *Glossary of Soil Science Terms*, Madison, WI.

CAROLYN G. OLSON

IMBRICATION

Imbrication is the orientation of an assembly of rocks or pebbles in one predominant direction (usually upstream), as a result of flow movement. Imbrication is common on gravel beaches, on bars and outwash fans in braided rivers, and glacial tills. Additionally, imbrication is often a key for interpreting facies. The term imbrication also refers to the near parallel overlapping and orientation of a series of lesser thrust faults, directed towards the source of stress.

Further reading

Byrne, J.V. (1963) Variations in fluvial gravel imbrication, *Journal of Sedimentary Petrology* 33, 467–469.

STEVE WARD

INHERITANCE

Landforms and landscapes have different lifetimes, which generally increase proportionally to the size of a landform. Therefore, many small landforms are formed and destroyed relatively quickly, over timescales of 10^3–10^4 yr, especially if the substratum is soft. By contrast, landforms built of hard, resistant rocks such as certain INSELBERGS could be very durable and may survive for much longer, over timescales of 10^5–10^7 yr. Likewise, regional landscapes such as extensive PLANATION SURFACES may have considerable lifetimes, measurable in millions of years. With such protracted histories, the bearing of present-day environmental conditions on their appearance and development is limited, and inheritance of the past becomes crucial.

Inherited landscapes are those that were formed in the past, under influences of a different external environment and/or within a tectonic regime different from the recent one, and have retained their characteristic features up to now, because later processes have proved to be inadequate to obliterate earlier sets of type landforms. Hence, there is no universal, worldwide applicable boundary between 'inherited' and 'contemporary'. In high latitudes of northern Europe and North America everything that predates Pleistocene continental glaciation is usually considered 'inherited', so that inherited landscapes are older than 1–1.5 Ma. By contrast, in Australia and South America some landscapes are likely to be inherited from the period preceding the breakup of Gondwana in the Mesozoic (Ollier 1991; Twidale 1994). In Central Europe, in turn, major change in the course of geomorphic evolution is often associated with growing tectonic instability from the Palaeogene/Neogene boundary onwards. Therefore, 'inherited' landforms would be those which originated within a general regime of tectonic quiescence and are older than ~20 Ma. However, in the specific context of Holocene morphogenesis, Pleistocene periglacial landscapes will also be regarded as inherited. It has to be emphasized that, strictly speaking, each landscape is recent in the sense that current processes do some action on it; yet the rate of change could be so negligible that the origin of gross features of such a landscape can be traced well back beyond the present day. In other situations, inherited forms may occur very close to relief features of recent origin, forming a palimpsest of landforms of contrasting origins and ages (Starkel 1987; Brunsden 1993a). In the case of the Fennoscandian Shield, where exhumed elements play a significant part, the range of ages of inherited landscapes spans from the Precambrian through the Jurassic, Cretaceous, late Tertiary up to the Pleistocene (Lidmar-Bergström 1995).

The phenomenon of landform and landscape inheritance has been recognized in geomorphology for a long time, though the very term 'inheritance' has not always been used. A German geomorphologist, S. Passarge (1919), made a distinction between *Vorzeitformen* (i.e. inherited) and *Jetztzeitformen* (i.e. moulded at present). Realization that the history of environmental change is very complex itself has led to the concept of 'generations of relief' (Büdel 1977). Accordingly, landscapes observed nowadays comprise units inherited from various epochs of the past, each recognizable through its characteristic set of landforms.

Inherited landscapes should not be equated with persistent landscapes as elaborated by Brunsden (1993b). The latter are landscapes which maintain their principal characteristics over time, and they do this for two main reasons. The first is that they have undergone very little change since their period of origin. The second is that persistent characteristics are maintained either because they are renewed or because they are uniformly changed. Thus while they maintain their original morphology, the currently observed landforms are not necessarily old. Therefore 'steady-state' landscapes are persistent. Zones of high relief may be maintained, for example, by erosion being compensated by isostatic adjustment. Hence, while every inherited landscape is persistent, persistent landscapes are not themselves necessarily inherited from the distant past.

Evidence for inheritance of land surfaces comes from three main sources. The most convincing one is offered by the occurrence of sediments of known age within denudational landscapes; these parts of landscapes are then at least contemporaneous with the sediments. Remnants of weathering mantles, including DURICRUSTS, are essentially of the same significance as the sediments, although the age of weathering residuals is much more difficult to establish. The third line of evidence is provided by the landforms themselves. In many cases there exists a sharp boundary between landscape units which helps to differentiate between older (inherited) and younger landforms. Examples include glacial U-shaped valleys breaching former watershed ridges, antecedent valleys cut into older surfaces and sea cliffs truncating subaerial landforms of long geomorphic history.

The survival of ancient landscapes means little geomorphological change, and these low rates of long-term denudation may be causally linked to one or more factors (Twidale 1999; Migoń and Goudie 2001). Subsiding areas retain their inherited geomorphological features for longer than areas subjected to surface uplift because erosional dissection and water divide lowering within the former is less effective. Moreover, as dissection proceeds and more rock mass is eroded, isostatic recovery (see ISOSTASY) becomes increasingly important, promoting further erosion and destruction of any remnants of old surfaces. However, rapid surface uplift with limited dissection seems capable of elevating a palaeosurface without transforming it into an all-slope topography. An inherited landscape will then be present at a high elevation.

Further reasons for inheritance may include climatic conditions unfavourable to rapid progress of denudation, high bedrock resistance, distance from base level or other lines of active erosion, and the protective role of durable duricrust blankets. In formerly glaciated countries the protective role of cold-based ice or the location within ice divides may play an additional part. A separate category is temporary burial and later re-exposure of an ancient landscape.

References

Brunsden, D. (1993a) Barriers to geomorphological change, in D.S.G. Thomas and R.J. Allison (eds) *Landscape Sensitivity*, 7–12, Chichester: Wiley.

——(1993b) The persistence of landforms, *Zeitschrift für Geomorphologie N.F. Supplementband* 93, 13–28.

Büdel, J. (1977) *Klima-Geomorphologie*, Stuttgart: Borntraeger.

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

Migoń, P. and Goudie, A.S. (2001) Inherited landscapes of Britain – possible reasons for survival, *Zeitschrift für Geomorphologie* 45(4), 417–441.

Ollier, C.D. (1991) *Ancient Landforms*, London: Belhaven.

Passarge, S. (1919) Die Vorzeitformen der deutschen Mittelgebirgslandschaften, *Petermann's Geographische Mitteillungen* 65, 41–46.

Starkel, L. (1987) The role of inherited forms in the present-day relief of the Polish Carpathians, in V. Gardiner (ed.) *International Geomorphology 1986*, Part II, 1,033–1,045, Chichester: Wiley.

Twidale, C.R. (1994) Gondwanan (Late Jurassic and Cretaceous) palaeosurfaces of the Australian Craton, *Palaeogeography, Palaeoclimatology, Palaeoecology* 112, 157–186.

——(1999) Landforms ancient and recent: the paradox, *Geografiska Annaler* 81A, 431–441.

SEE ALSO: climato-genetic geomorphology

PIOTR MIGOŃ

INITIATION OF MOTION

The initiation of sediment transport (also known as incipient motion or the critical condition) describes when one or more particles are moved from a stationary bed. In river or air flows, the force balance for a particle resting on the surface can be resolved into three components: a drag force acting parallel to the bed, known as the shear stress (τ), an upward lift force produced by turbulence and the downward frictional force caused by the particle's weight. In 1936, Shields showed that the onset of particle motion is defined as $\tau_c^* = \tau_c/(\rho_s - \rho)gD$ where τ_c^* is the dimensionless shear stress at the critical condition, τ_c is the tangential fluid shear stress at the boundary as grain movement begins and ρ_s, ρ, g and D are the sediment density, fluid density, acceleration due to gravity and grain diameter at the bed surface, respectively. Using a series of laboratory flume experiments with sediment of uniform grain size, Shields showed that the value of τ_c^* lies within a narrow range for hydraulically rough and turbulent flow (mean $\tau_c^* = 0.046$ when corrected for sidewall effects and form drag). Consequently, the Shields equation stated above simplifies to $\tau_c \propto D$ which implies that a stronger fluid force will move coarser particles. Some authors refer to this finding as the principle of 'selective entrainment'.

Although intuitively attractive, it has now been shown that there is a frequency distribution of τ_c^* values for any particular grain size. This is because the three-dimensional forces of turbulent shear stress that impact on the bed vary through space and time and the potential for movement of a particular grain is strongly dependent on the physical arrangement of the bed surface (e.g. grain pivoting angle, degree of packing and sorting, mixture of grain shapes). The variation in τ_c^* is most marked over coarse-grained and hydraulically rough boundaries where different particle sizes protrude or are 'hidden' behind and below other particles on the bed surface. Researchers have shown that τ_c^* decreases with an increase in relative particle size (i.e. the ratio between the size of the surface particle and the ambient grains). So long as there is an overrepresentation of coarse sediment on the bed surface, sediment entrainment may achieve what is termed 'equal mobility' whereby all particles in the surface may be moved regardless of their absolute size or weight and the BEDLOAD grain size distribution matches that of the bed subsurface.

Further reading

Buffington, J.M. and Montgomery, D.R. (1997) A systematic analysis of eight decades of incipient motion studies, with special reference to gravel-bedded rivers, *Water Resources Research* 33, 1,993–2,029.

Kennedy, J.F. (1995) The Albert Shields story, *Journal of Hydraulic Engineering*, American Society of Civil Engineers 121, 766–771.

Lavelle, J.W. and Mofjeld, H.O. (1987) Do critical stresses for incipient motion and erosion really exist? *Journal of Hydraulic Engineering*, American Society of Civil Engineers 113, 370–385.

SEE ALSO: downstream fining; fluvial armour; mobile bed

PHILIP J. ASHWORTH

INLAND DELTA

The term was coined by cartographers in the late nineteenth century to describe a morphologically unusual reach of the Niger River in the interior of west Africa (Mali), which they termed the Niger Inland Delta. Here, in the vicinity of the confluence (see CONFLUENCE, CHANNEL AND RIVER JUNCTION) of the Niger and Bani Rivers, the two rivers divide to form a network of anastomosed (see ANABRANCHING AND ANASTOMOSING RIVER) channels spread over a FLOODPLAIN approximately 100 km wide and 200 km long. The channels terminate in Lakes Debo and Korientze, which are ponded against aeolian dunes on the southern edge of the Sahel. A single channel forms beyond the dunefield. The adjective 'inland' was added to distinguish this feature from the delta (see RIVER DELTA) at the mouth of the Niger. A form of the term was also applied to the terminus of the Okavango River in the endoreic Kalahari Basin of central southern Africa (Botswana). After meandering through a 10 to 15 km wide, 100 km long corridor known as the Panhandle, the Okavango River abruptly divides into a number of radial, distributary channels extending over a floodplain some 50,000 km^2 in area, known as the Okavango Delta. The adjective 'inland' was omitted in this case because the Okavango River, unlike the Niger, does not reach the ocean, so there is no risk of confusion. A geomorphologically similar feature is also developed on the White Nile River (Bahr el Jebel) in Sudan, where the river divides into several distributary channels forming a triangle 520 km long and 150 km along its base. The term 'delta' has not been applied in this case, however, and the feature is known by its Arabic name, the Sudd.

The Niger Inland Delta, Okavango Delta and the Sudd have several features in common. They occur in actively subsiding half-grabens, possibly related to incipient rifting. The multiple channels of these inland deltas form in response to the loss of confinement of flow as the rivers enter the grabens. They have relatively low gradients: 3 cm km^{-1} for the Niger, 10 cm km^{-1} for the Sudd and 28 cm km^{-1} for the Okavango. River discharge is seasonal and the flood wave moves slowly down the floodplain, taking three months to cross the Niger Inland Delta, four months to cross the Sudd and five months to cross the Okavango Delta. There is considerable evapotranspirational loss of water – 98 per cent in the case of the Okavango.

Whilst the channel geometry of these inland deltas is superficially similar to those on deltas at river mouths, the sedimentation processes are fundamentally different. In river deltas, sedimentation occurs primarily on the delta front as the distributary channels lose their ability to transport sediment on entering standing-water bodies. River deltas therefore build outwards into water bodies. In contrast, on inland deltas sediment deposition occurs primarily as a result of flood water spilling from channels and spreading laterally as sheet flow. Inland deltas therefore aggrade vertically. Local AGGRADATION around channels results in instability, causing AVULSION which gives rise to new channels. In this way, sediment is spread uniformly across the delta surface. This sedimentary style is more akin to that occurring on ALLUVIAL FANs, and these inland deltas constitute a variety of low gradient alluvial fan.

Further reading

Howell, P., Lock, M. and Cobb, S. (1988) *The Jonglei Canal*, Cambridge: Cambridge University Press.

McCarthy, T.S. (1993) The great inland deltas of Africa, *Journal of African Earth Sciences* 17, 275–291.

Makaske, B. (1998) *Anastomosing Rivers: Forms, Process and Sediments*, Faculteit Ruimtelijke Wetenschappen, Universiteit Utrecht.

Stanistreet, I.G. and McCarthy, T.S. (1993) The Okavango fan and the classification of subaerial fan systems, *Sedimentary Geology* 85, 115–133.

SEE ALSO: alluvial fan; megafan; river delta

T.S. McCARTHY

INSELBERG

Inselberg is a descriptive term, derived from German (*inselberg* literally means 'island hill') and adopted in English, used to describe a hill which stands in isolation and rises sharply above the level of the surrounding plain. It was first coined by W. Bornhardt, a German naturalist travelling in East Africa at the turn of the nineteenth century, to emphasize visual similarities between steep-sided hills dotting an otherwise flat savanna and islands rising from the sea. Inselbergs often tend to occur in groups, to form inselberg landscapes.

It is generally accepted that the term applies to a hill produced by lowering of the surface around it. Therefore volcanoes and small tectonic horsts are normally not called inselbergs (Kesel 1973). It was also noted that the term is rather infrequently used to describe isolated hills built of flat-lying sedimentary rocks such as MESA or BUTTE, hence the impression arises that inselberg landscapes are restricted to basement rock areas. There have been attempts to make the definition more strict by using certain quantitative constraints for inselbergs. These included minimal height of 15 m, length to width ratio not exceeding 4 : 1, or minimal distance to the nearest neighbour of 0.8 km (Faniran 1974). None of these proposals has been universally accepted and there is quite a degree of freedom in deciding which hill is to be called an inselberg, which makes comparative analyses of inselberg landscapes difficult.

Visual differences provide the basis for classification of inselbergs, most readily applicable to granite hills. Three main types are distinguished (Thomas 1978; Twidale 1981). Domed inselbergs have slopes convex-upward and are built of massive, poorly jointed rock, with little REGOLITH at the surface. They are called BORNHARDTS, but it is important to note that the two terms are not equivalents. Hills having all the characteristics of bornhardts listed above can also occur within hilly and mountainous terrains, and hence are not by any means inselbergs. Examples include 'sugarloaf' hills in Rio de Janeiro, or granite domes in the Yosemite National Park in the USA. Castellated inselbergs are built of well-jointed but bedrock-rooted rock and their detailed morphology consists of towers, pillars and walls, separated by joint-aligned avenues. Boulder inselbergs are composed of loose boulders, chaotically lying one upon another. They typically form through advanced degradation of a domed or castellated inselberg. In addition, some inselbergs, especially in metamorphic rocks, may be conical, and there are also hills built entirely of SAPROLITE.

Inselbergs tend to be built of hard, resistant igneous and metamorphic rocks such as granite or gneiss. They are particularly common in coarse-grained and poorly jointed granites, rich in potassium feldspar (Pye *et al.* 1986). They are also known from gabbro, syenite, rlyolite and migmatite. However, there are relatively few instances where the base of the hillslope coincides with a lithological boundary, nor are differences in mineralogical composition between the hill and the plain significant (Brook 1978). The majority of inselbergs show joint- rather than lithological control. Bedrock to build an inselberg is typically massive, with few open fractures and the predominance of tight SHEETING joints. The latter are often arranged concentrically to form a structural dome, the outline of which is followed by the topography. It is generally assumed, although usually not proved, that joint density around a hill is higher than within the hill, and the existence of an inselberg reflects primary variations in the degree of rock mass fracturing. Another manifestation of joint control is seen in plan. Outlines of inselbergs frequently follow master joints (Twidale 1982; Selby 1982).

Notwithstanding the impression that inselbergs are only supported by igneous rocks, spectacular inselberg assemblages can also be built of sedimentary rocks, especially of massive sandstone and conglomerate. Ayers Rock (Uluru) in central Australia, one of the most famous inselbergs worldwide, is made of steeply dipping arkosic sandstone, whereas the adjacent Olgas are composed of massive conglomerate. Further examples of isolated sandstone towers have been reported from certain arid areas, such as the Sahara Desert in Niger and Mali, or the deserts of south-west Jordan. Interestingly, sandstone inselbergs in these areas are not associated with any actively retreating escarpments but rather have originated due to advanced dissection of a former plateau.

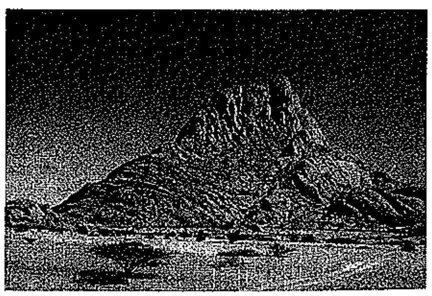

Plate 65 The granite inselberg of Sptizkoppe in the Namib Desert is built of extremely massive – and therefore most resistant – granite, which accounts for its impressive height of more than 600 m

The origin of inselbergs had been a matter of hot debate, but it is now accepted that they may form in various ways, and contrasting evolutionary pathways may produce landforms looking superficially very similar. There are at least three theories present in the literature. Perhaps the most universally accepted one holds that inselbergs are products of two-stage development involving differential deep weathering in the first phase and stripping of the weathering mantle in the second one, which leaves an unweathered rock compartment exposed at the surface (Thomas 1965). Reasons for survival of such compartments include significantly wider spacing of joints, the presence of primary poorly fractured mass, enrichment in quartz and/or potassium feldspar, or petrological differences. Validity of the two-stage hypothesis has been confirmed in deep excavations and quarries in equatorial Africa, where massive hill-like features up to 50 m high have been found within a thick mantle of decomposed bedrock. In reality, the average thickness of weathering mantles seems insufficient to account for the height of many inselbergs, which may exceed 200–300 m. Therefore their exposure is more likely to have been accomplished in many stages of weathering and stripping (Twidale and Bourne 1975). Minor landforms on hillslopes such as flared slopes and platforms are considered as the evidence of multiphase exposure.

Another hypothesis invokes scarp retreat across unweathered, but possibly differentially jointed, bedrock and links the origin of inselbergs with the cyclic development of relief and retreat of big escarpments separating denudational landscapes of various ages (King 1949). Clustering of inselbergs in front of major escarpments might validate this theory, but other people argue for an important role of deep weathering in the development of scarps and inselbergs too (Bremer 1993).

The massive nature of many prominent inselbergs found in the semi-arid and arid zone, and the occurrence of more jointed compartments around their flanks, suggest that exposure and growth in height may not necessarily be associated with deep weathering. Long-term lowering of differentially jointed bedrock probably accounts for the origin of the spectacular inselbergs of the Namib Desert (Selby 1982) and may be applicable elsewhere. In the same way, some minor granite intrusions may have been exposed as inselbergs as the surrounding less resistant schist has been completely eroded away.

Inselbergs, once exposed or isolated, undergo further development and are the scene of competing processes of continuous growth and destruction. Many authors emphasize that inselbergs are very durable, long-lived landforms because their surfaces shed rainwater and remain dry, being therefore immune to chemical weathering. At the same time, because they are built of poorly jointed bedrock, they are resistant against physical weathering too. Fast runoff from the slopes of an inselberg provides additional moisture to its footslopes, so the rate and intensity of weathering are enhanced. Episodic removal of regolith cover from around the inselberg may then result in the increase of its height, providing the lowering of its summit proceeds at a slower rate. In Australia, it is argued that inselbergs have been rising for millions of years and their top surfaces may date back to the Mesozoic (Twidale and Bourne 1975; Twidale 1978).

How inselbergs are reduced in height and extent depends on their jointing patterns. Massive domed inselbergs are subject to mega-exfoliation due to pressure release and opening of sheeting joints. Individual slabs are separated from the underlying rock mass and fall or slide off the slopes, forming big debris cones mantling lower slopes. Ongoing exfoliation will gradually reduce the surface area of an inselberg. Rock fall is also typical for slope development of sandstone inselbergs in deserts. In orthogonal patterns, vertical joints open too, topples occur, and the summit part assumes ruiniform relief. Minor weathering features play an important part in the development of inselbergs (Watson and Pye 1985). Selective weathering along fractures and growth of TAFONI reduces rock mass strength and facilitates mass movement, whereas horizontal surfaces are destroyed by enlargement of WEATHERING PITS. Caves and massive overhangs are frequently reported from granite inselbergs and develop either through mechanical widening of fractures, preferential weathering along sheeting joints, or chaotic accumulation of big boulders on lower slopes.

The discussion about the origin of inselbergs has a direct bearing on the issue of their significance in geomorphology, especially in CLIMATIC GEOMORPHOLOGY and CLIMATO-GENETIC GEOMORPHOLOGY. Two positions emerge from the literature. One holds that inselbergs occur all around the world and cannot be considered as indicators of environments, present or past (Kesel 1973; King 1975). King argued that the process of scarp retreat is climate-independent, hence inselbergs are not dependent on climatic conditions either. By contrast, a German school of geomorphology has maintained that inselbergs are specific products of landscape development in the seasonally humid tropics and develop through deep weathering and stripping, which are the processes acting at their highest efficacy in this zone. Consequently, inselbergs present in middle or high latitudes, or in arid areas, would be relict landforms, inherited from the geological past.

Although it is probably true that humid tropical environments with ubiquitous deep weathering favour the development of inselbergs, a claim that inselbergs are by definition 'tropical' landforms is most likely wrong. Inselbergs are present in many desert areas of the world, including the long-lived ones such as the Namib, and in many evidence for inheritance from previously humid conditions is clearly lacking. Nevertheless, the origin of those present in central and northern Europe, as well as in North America, is usually traced back to the Early Tertiary, when climate was warmer and wetter, and deep weathering was widespread. It is worth noting that even if these inselbergs are indeed ancient Tertiary features, it does not mean that tropical climate was essential for their origin.

Inselberg landscapes have been reported from all around the world, including parts of Antarctica, but they are probably most widespread in Africa, within extensive tracts of crystalline rock terrain. Classic examples are known from Nigerian savannas, East African plains in Kenya and Tanzania, Zimbabwe, South Africa, Namib Desert and Angola. Namibian inselbergs, such as the almost 700-m high Spitzkoppe, belong to the highest in the world. There are also examples from the Sahara and its southern margin, from Sudan, Niger and Libya. Inselbergs are also common in Australia, especially in the central and western part of the continent. Further examples are known from the Indian Peninsula and basement areas of South America. Numerous, purportedly inherited, inselberg landscapes have been described in Europe, including Germany, the Czech Republic, Poland, Hungary and Scandinavia.

References

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

Brook, G.A. (1978) A new approach to the study of inselberg landscapes, *Zeitschrift für Geomorphologie Supplementband* N.F. 31, 138–160.

Faniran, A. (1974) Nearest-neighbour analysis of inter-inselberg distance: a case study of inselbergs of south-western Nigeria, *Zeitschrift für Geomorphologie Supplementband* N.F. 20, 150–167.

Kesel, R.H. (1973) Inselberg landform elements: definition and synthesis, *Revue Géomorphologie Dynamique* 22, 97–108.

King, L.C. (1949) A theory of bornhardts, *Geographical Journal* 113, 83–87.

——(1975) Bornhardt landforms and what they teach, *Zeitschrift für Geomorphologie* 19, 299–318.

Pye, K., Goudie, A.S. and Watson, A. (1986) Petrological influence on differential weathering and inselberg development in the Kora area of Central Kenya, *Earth Surface Processes and Landforms* 11, 41–52.

Selby, M.J. (1982) Form and origin of some bornhardts of the Namib Desert, *Zeitschrift für Geomorphologie Supplementband* N.F. 26, 1–15.

Thomas, M.F. (1965) Some aspects of the geomorphology of domes and tors in Nigeria, *Zeitschrift für Geomorphologie Supplementband* N.F. 9, 63–82.

——(1978) The study of inselbergs, *Zeitschrift für Geomorphologie Supplementband* N.F. 31, 1–41.

Twidale, C.R. (1978) On the origin of Ayers Rock, central Australia, *Zeitschrift für Geomorphologie Supplementband* N.F. 31, 177–206.

——(1981) Granite inselbergs: domed, block-strewn and castellated, *Geographical Journal* 147, 54–71.

——(1982) *Granite Landforms*, Amsterdam: Elsevier.

Twidale, C.R. and Bourne, J.A. (1975) Episodic exposure of inselbergs, *Geological Society of America Bulletin* 86, 1,473–1,481.

Watson, A. and Pye, K. (1985) Pseudokarstic micro-relief and other weathering features on the Mswati Granite (Swaziland), *Zeitschrift für Geomorphologie Supplementband* N.F. 29, 285–300.

PIOTR MIGOŃ

INSOLATION WEATHERING

Insolation weathering (thermoclasty or thermal stress fatigue) is the rupturing of rocks and minerals primarily as a result of large daily temperature changes in dry environments which lead to temperature gradients within the rock mass. Fires can operate in a similar way, though the temperature extremes are greater (see FIRE). Areas that are heated expand relative to the cooler portions of the rock and stresses are thereby set up.

In igneous rocks, which contain many different types of mineral (polymineralcy) with different coefficients and directions of expansion, such stresses are enhanced. Moreover, the varying colours of minerals exposed at the surface (polychromacy) will cause differential heating and cooling.

Daily temperature cycles under desert conditions may exceed 50 °C, and during the heat of the day rock surfaces may occasionally exceed a temperature of 80 °C. However, rapid cooling takes place at night, creating, it has been thought, high tensile stresses in the rock. Desert travellers, like David Livingstone, have claimed to hear rocks splitting with sounds like pistol shots in the cool evening air – certainly, split rocks are evident on many desert surfaces. Insolation was a mechanism that found considerable favour with pioneer desert geomorphologists (e.g. Hume 1925; Walther 1997).

At first sight the process of insolation weathering seems a compelling and attractive mechanism of rock disintegration. However, doubt has been cast upon its effectiveness upon a variety of grounds (Schattner 1961). The most persuasive basis for doubting its power was provided by early experimental work in the laboratory by geomorphologists like Blackwelder, Griggs and Tarr. They all found that simulated insolation produced no discernable disintegration of dry rock, but that when water was used in the cooling phase of a weathering cycle disintegration was evident. This highlighted the importance of the presence of water. Likewise, studies of ancient buildings and monuments in dry parts of North Africa and Arabia showed very little sign of decay except in areas, for example close to the Nile, where moisture was present (Barton 1916). Indeed, there are many situations where there is moisture in deserts (e.g. where there is fog, dew and groundwater seepage). When water combines chemically with the more susceptible minerals in a rock they may swell, producing a sufficient increase in volume to cause the outer layers of rock to be lifted off as concentric shells, a process called EXFOLIATION. Thus, some of the weathering

that used to be attributed to insolation may now be attributed to chemical changes produced by moisture, including HYDRATION.

However, the importance of insolation cannot be dismissed entirely. The early experimental work had grave limitations: the blocks used were very small and unconfined, the temperature cycles were unrealistic and only a limited range of rock types was used (Rice 1976). Moreover, engineering and ceramic studies have shown that a threshold value for thermal shock approximates to a rate of temperature change of $2\,°\mathrm{C\,min}^{-1}$. Datalogger studies show that such rates can occur, not least in polar regions (Hall and André 2001). In addition, consideration of fracture patterns observed on rock in cold, dry environments appears to show very similar forms to those produced in thermal shock experiments in the laboratory (Hall 1999). Finally, because of the temperature response of calcite crystals, marble seems to be especially prone to thermal degradation (Royer-Carfagni 1999)

References

Barton, D.C. (1916) Notes on the disintegration of granite in Egypt, *Journal of Geology* 24, 382–393.

Hall, K. (1999) The role of thermal stress fatigue in the breakdown of rock in cold regions, *Geomorphology* 31, 47–63.

Hall, K. and André M.-F. (2001) New insights into rock weathering from high-frequency rock temperature data: an Antarctic study of weathering by thermal stress, *Geomorphology* 41, 23–35.

Hume, W.F. (1925) *Geology in Egypt*, Vol. 1, Cairo: Government Press.

Rice, A. (1976) Insolation warmed over, *Geology* 4, 61–62.

Royer-Carfagni, G.F. (1999) On the thermal degradation of marble, *International Journal of Rock Mechanics and Mining Sciences* 36, 119–126.

Schattner, I. (1961) Weathering phenomena in the crystalline of the Sinai in the light of current notions, *Bulletin Research Council of Israel* 106, 247–266.

Walther, J. (1997) *The Law of Desert Formation – Present and Past* (translated by G. Meyer and edited by E. Gischler and K.W. Glennie), University of Miami, Geological Milestones, Vol. 4.

A.S. GOUDIE

INTEGRATED COASTAL MANAGEMENT

Coastal geomorphology and associated coastal resources are subject to increasing pressure from human impact. In a global context, this is significant for two reasons. First, the majority of the world's population currently lives near the coast and the proportion of coastal dwellers is projected to increase in the future. The world's coastal population has been variously estimated as over half living within 60 km of the coast (UNCED 1992: 17.3); 1.2 billion within 100 km of the coast (Nicholls and Small 2002, based on 1990 data); or 3.2 billion within 200 km and two-thirds within 400 km (Hinrichsen 1998). Second, humans have a high dependence on coastal resources. According to key researchers in resource economics (Costanza *et al.* 1997) the coastal biome currently contributes over 40 per cent of the total global flow value of ecosystem services.

These global population pressures and human dependence on the coast require appropriate management strategies which recognize the dynamic nature of the coast at various geomorphic timescales and across different spatial dimensions. While longer timescales may be less problematic from a management perspective, the identification of rapid geomorphic change may be difficult to separate from human-induced change. For example, coastal wetlands may be subject to human impacts of reclamation, development or subsidence through groundwater withdrawal acting simultaneously with natural processes of wetland loss such as local relative sea-level change or sediment compaction. It is also important to recognize that local coastal impacts may have broader spatial linkages to marine or land-based processes. For example, coastal pollution, erosion or accretion near a river mouth may be attributed to poor catchment management practices rather than localized coastal processes.

For these reasons, there has been a recognition that the coast, along with other types of environment, needs to be managed in a holistic rather than a sectoral manner. This has given rise to an increasing acceptance of integrated resource management in general, and more specifically to integrated coastal management (ICM). Although the concept of coastal management has been around for more than thirty years, particularly with the introduction of coastal legislation to the United States in 1972, it was the United Nations Conference on Environment and Development (UNCED) (also known as the 'Earth Summit') held in Rio de Janeiro in 1992 that created an international push for the adoption of 'integrated' coastal management with an agreement that coastal states should 'commit themselves to integrated management and sustainable development of coastal areas and the marine environment under their national jurisdiction' (UNCED 1992: Agenda 21: 17.5).

Similar objectives are contained within the United Nations Framework Convention on Climate Change (1992) which outlines the need to develop integrated plans for coastal management. In the following year (1993), the Council of the Organisation for Economic Cooperation and Development and the first World Coastal Conference adopted and produced guidelines for the integrated management of coastal resources. These *inter alia* required that nations developing their ICM should take into account the traditional, cultural and historical perspectives and conflicting interests and uses (IPCC 1994). In order to define ICM it is useful to draw on earlier definitions from the IPCC (1994: 40), Cicin-Sain and Knecht (1998: 39) and also from the Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP 1996: 2).

Integrated Coastal Management is a continuous and dynamic process incorporating feedback loops which aims to manage human use of coastal resources in a sustainable manner by adopting a holistic and integrative approach between terrestrial and marine environments; levels and sectors of government; government and community; science and management; and sectors of the economy.

It is important to realize that although there are a number of definitions of ICM, there is still some confusion over the use of other related coastal management terms and definitions. Cicin-Sain and Knecht (1998) and Burbridge (1999) note that there has been a major change in emphasis away from coastal 'zone' or 'area' management towards 'integrated' coastal management. Cicin-Sain and Knecht argue that the terms integrated coastal zone management (ICZM), integrated marine and coastal area management and integrated coastal management (ICM) all refer to the same concept and they adopt ICM for reasons of consistency and simplicity. Others still use the term ICZM (see Salomons *et al.* 1999) although some authors in the same volume (Burbridge 1999; Harvey 1999) prefer the term ICM. In a major review of the coastal management literature Sorensen (1997) distinguishes between ICM as a concept or field of study and ICZM as a programme which has the task of defining the boundaries of the coastal 'zone'. On balance it appears that although the use of the term 'zone' was originally intended to be flexible, it can also be interpreted as prescriptive if the identification of boundary conditions mitigate against the need to integrate across them. For this reason, the use of ICM is becoming more acceptable and common in the literature.

There is now a global trend toward a more integrated approach for coastal management which incorporates linkages between activities in coastal lands and waters. A decade ago, the Earth Summit (UNCED 1992) recognized the need for a new approach to marine and coastal area management developed at the national, subregional, regional and global levels. It also commented that any new approach to coastal management should be integrated in content, and precautionary and anticipatory in its scope. Subsequently, there have been various international attempts to develop guidelines for ICM, stressing the importance of strengthening and harmonizing cross-sectoral management. While there are different approaches for achieving ICM, most agree that horizontal and vertical integration and co-ordination must be part of any attempt to achieve ICM.

Cicin-Sain and Knecht (1998) suggest that the rationale for an integrated approach is first to examine the effects of ocean and coastal use, as well as activities further inland, on ocean and coastal environments; and second to examine the effects that ocean and coastal users can have on one another. The World Coast Conference puts it simply as the need for co-operation between all responsible actors involved in coastal management (IPCC 1994: 25). The key elements of integration in coastal management can be defined as follows:

- Intergovernmental integration (vertical integration) between different levels of government such as national, provincial or state and local governments;
- Intersectoral integration (horizontal integration) between different government sectors: such as industry, conservation, recreation, tourism, beach protection and integration of policies between different sectors of the economy;
- Community integration with government producing effective community participation and involvement in coastal management;

- Spatial integration between management of the land, ocean and coast;
- Integration between science and management particularly between different disciplines; scientists and managers; including economic, technical and legal approaches to coastal management;
- International integration between nations on trans-boundary coastal management issues.

There are a number of requirements for the success of ICM. There is a need for a long-term strategy with clearly defined national objectives and guiding principles for coastal managers. It is also necessary to have a defined authority with responsibility for the strategy along with monitoring of key performance indicators and most importantly there is a need for the political will for implementation of ICM. Various conceptual models (e.g. Cicin-Sain and Knecht 1998: 58) of the different stages in the ICM process all emphasize the cyclical nature of the process and the need for continual re-evaluation. The implication of this is a need to proceed through all stages at least once before the success of ICM can be properly evaluated. This is likely to take a number of years.

Since the Earth Summit and the World Coast Conference, ICM has been adopted by many nations with coastal management programmes and associated legislation. Cicin-Sain and Knecht (1998) conclude from their cross-national ICM survey in 1996 that there were approximately 150 national ICM efforts globally, including the following countries: United States, United Kingdom, Belize, Brazil, Costa Rica, Ecuador, Sri Lanka, Turkey, Australia, Canada, Italy, China, Mexico, Nigeria, Venezuela and Pohnpei State (Federated States of Micronesia). Subsequently, Sorensen has created a global data base (www.uhi.umb.edu) which, in 2000, contained a total of 385 'ICM efforts' comprising 250 in 87 countries, plus 100 in the United States and 35 internationally. However, care is needed in interpreting Sorensen's data on ICM efforts which contain a mixture of programmes, policy statements and feasibility studies.

Notwithstanding the increasing global number of ICM efforts, there is a paucity of data on their success; for example, Cicin-Sain and Knecht's comment that their 1996 survey results provided scanty evidence on the extent of ICM implementation and effectiveness. They concluded that it was very difficult to produce a model of successful ICM because there was a lack of objective evaluation data for any of the ICM examples they described (Cicin-Sain and Knecht 1998: 294). Sorensen (1997) comments that there is uncertainty in our knowledge about the important implementation phase of ICM programmes which comes after the adoption phase of various plans and policies. Burbridge (1997) makes a distinction between developing ICM initiatives and assessing their success in meeting stated goals and he suggests that there are not many good examples of fully developed ICM strategies, plans or practices that extend beyond a local or problem-specific level (Burbridge 1997: 181). It is also important to note that there are significant socio-economic and political differences between coastal nations which need to be considered in assessing ICM achievements.

Thus the concept of ICM is internationally accepted as an appropriate method for sustainable management of coastal resources. The concept has evolved from a realization that a sectoral approach to coastal management is fundamentally flawed. There is now a global proliferation of what has been loosely termed ICM efforts although there is scope for more rigorous survey of these. Finally, it is clear that the international acceptance of ICM as an approach is not matched by definable criteria or models to judge its success or best practice.

References

Burbridge, P.R. (1997) A generic framework for measuring success in Integrated Coastal Management, *Ocean and Coastal Management* 37(2), 175–189.

——(1999) The challenge of demonstrating the socio-economic benefits of Integrated Coastal Management, in W. Salomons, R.K. Turner, L.D. de Lacerda and S. Ramachandran (eds) *Perspectives on Integrated Coastal Zone Management*, 35–53, Berlin: Springer-Verlag.

Cicin-Sain, B. and Knecht, R.W. (1998) *Integrated Coastal and Ocean Management: Concepts and Practices*, Washington, DC: Island Press.

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Linburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. and van den Belt, M. (1997) The value of the world's ecosystem services and natural capital, *Nature* 387, 253–260.

GESAMP (IMO/FAO/UNESCO-IOC/WMO/WHO/IAEA/UN/UNEP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) (1996) *The Contributions of Science to Coastal Zone Management*, GESAMP Reports and Studies 61.

Harvey, N. (1999) Australian Integrated Coastal Management: a case study of the Great Barrier Reef, in W. Salomons, R.K. Turner, L.D. de Lacerda and S. Ramachandran (eds) *Perspectives on Integrated Coastal Zone Management*, 279–296, Berlin: Springer-Verlag.

Hinrichsen, D. (1998) *Coastal Waters of the World: Trends, Threats and Strategies*, Washington, DC: Island Press.

Intergovernmental Panel on Climate Change (IPCC) (1994) *Preparing to Meet the Coastal Challenges of the twenty-first Century. Report of the World Coast Conference, 1–5 November 1993*, The Hague: Ministry of Transport, Public Works and Water Management.

Nicholls, R.J. and Small, C. (2002) Improved estimates of coastal population and exposure to hazards released, *EOS, Transactions, American Geophysical Union* 83(28), 301–305.

Salomons, W., Turner, R.K., de Lacerda, L.D. and Ramachandran, S. (eds) (1999) *Perspectives on Integrated Coastal Zone Management*, Berlin: Springer-Verlag.

Sorenson, J. (1997) National and international efforts at Integrated Coastal Zone Management: definitions, achievements, and lessons, *Coastal Management* 25, 3–41.

United Nations Conference on Environment and Development (UNCED) (1992) *Agenda 21 and the UNCED Proceedings*, New York: Oceania Publications.

United Nations Framework Convention on Climate Change (UNFCC) (1992).

NICK HARVEY

INTERDUNE

Interdune areas are the depressions between dunes. Interdune areas occur in a variety of shapes, which reflect the morphology of the dunes with which they are associated. Fields of crescentic dunes typically have ellipsoidal interdune areas that are elongate with the dune crestline. Extremely long interdune corridors occur with linear dunes, whereas more irregular, interconnected interdune areas accompany complicated dune shapes such as star dunes. There is a complete gradation from interdune flats that cover a greater portion of the field than the dunes, to interdune depressions between the lee and the stoss (upwind) slopes of adjacent dunes.

At least for transverse dunes, the formation of an interdune area can be visualized by watching a fixed point at the base of the stoss slope of a migrating dune. The interdune area begins at this erosional point and extends downwind as an interdune BOUNDING SURFACE as the stoss slope migrates. The surface continues to extend downwind until it is buried by the lee deposits of the next dune upwind, which has now migrated to the fixed point.

Although easy to visualize, the explanation for interdune areas remains controversial, and is intimately linked with the explanation of how dunes come to be regularly spaced. In one hypothesis, dune spacing is a function of the fluid dynamics established with airflow over dunes. With transverse dunes, there is flow separation at the dune brink, creating a separation cell in the immediate lee, and flow reattachment at a distance of a few dune heights from the brink (Walker and Nickling 2002). At the reattachment point, a new BOUNDARY LAYER is created. Wind speed and shear stress increase downwind within this new boundary layer, owing to the downward flux of momentum from the overlying, relatively high-speed flow. As long as the surface winds are accelerating, there is at least potential deflation of the interdune area (Kocurek *et al.* 1991). The boundary layer, however, is also expanding vertically, and at some point this expansion overwhelms the effects of the momentum flux, and wind speed and shear stress within the boundary should decrease. At that point, deposition (and initiation of a new dune) can begin. In this way, the spacing of the next dune downwind and the length of the interdune area is defined by the fluid dynamics established by the upwind dune. Field data, however, show that flow speed and shear stress reach a steady-state condition that favours sediment bypass, and not deceleration and deposition (Frank and Kocurek 1996). Adequate data do not yet exist to evaluate the role of turbulence, which may play a more significant role in the spacing of subaqueous dunes (Nelson *et al.* 1995).

In an alternative view, fields of dunes are viewed as self-organizing and the interdune areas exist by default. Computer simulations have modelled the formation of all major dune types as a function of the direction and duration of sand-transporting winds, and independent of the creation of an internal boundary layer (Werner 1995). In this hypothesis, dunes begin as small, randomly spaced collections of sand that merge and grow larger as they migrate, progressing to a steady-state condition where little change is possible because all the dunes are about the same size and, therefore, migrating at about the same speed. Initiation of sand dunes and their development to a steady-state field on Padre Island, Texas, strongly resemble the model (Kocurek *et al.* 1991). The interdune areas in this model are then simply the areas where the dunes are not (Werner and Kocurek 1999).

Regardless of their origin, interdune areas contain some of the most diverse features seen in dunefields. The first division is between those in which accumulation occurs, and those that are deflationary, in which older dune accumulations, reg or bedrock can be exposed. For deflationary interdune areas, or ones in which sediment bypass only occurs, the interdune flats essentially form a continuous surface over which the field of dunes migrate. Characteristic of dunefields where the water table controls the level of deflation (wet aeolian systems), corrugated surfaces develop on the interdune floor that reflect the stratification types of underlying dune accumulations.

Where sediment accumulation occurs on the interdune surface, these sediments can be classified as dry-, damp-, and wet-surface deposits (Kocurek 1981). Dry-surface, sandy accumulations are typically those from wind ripples, grainfall from the upwind dune and satellite dunes. Because the surface is dry and subject to deflation, these interdune accumulations within dry aeolian systems are usually restricted to interdune depressions that exist within the low-wind-speed separation cell. Extensive dry interdune flats, however, are common and these may contain concentrations of less-easily transported coarse grains, typically organized into granule ripples and zibar dunes. Sandy deposits are usually concentrated around sediment-trapping vegetation.

Where the interdune surface is wet or damp, typically because the water table is near the surface, a major division occurs between those that contain evaporate minerals and those that do not. Interdune areas in temperate climates (e.g. coastal fields) typically lack evaporites and the deposits are dominated by ADHESION structures, wrinkle marks, fluid-escape structures and subaqueous ripples, mud drapes, algal mats, channels and other features formed during ponding. In arid climates, wet- or damp-surface interdune areas are nearly always characterized by evaporites. These may occur as interdune SABKHAs, which show a characteristic crinkly sediment texture that results from deposition by salt ridges. Because interdune surfaces in wet aeolian systems have a high capillary water content or cementing evaporites, accumulation can occur on these flats in spite of flow acceleration within the interdune boundary layer, and commonly these interdune areas contain the finest sediment within the field.

References

Frank, A. and Kocurek, G. (1996) Toward a model of airflow on the lee side of aeolian dunes, *Geomorphology* 17, 47–54.
Kocurek, G. (1981) Significance of interdune deposits and bounding surfaces in aeolian dune sands, *Sedimentology* 28, 753–780.
Kocurek, G., Townsley, M., Yeh, E., Havholm, K. and Sweet, M.L. (1991) Dune and dunefield development on Padre Island, Texas, with implications for interdune deposition and water-table-controlled accumulation, *Journal of Sedimentary Petrology* 62, 622–635.
Nelson, J.M., Shreve, R.L., McLean, S.R. and Drake, T.G. (1995) Role of near-bed turbulence structure in bedload transport and bedform mechanics, *Water Resources Research* 31, 2,071–2,086.
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.
Werner, B. (1995) Eolian dunes: computer simulations and attractor interpretation, *Geology* 23, 1,107–1,110.
Werner, B. and Kocurek, G. (1999) Bedform spacing from defect dynamics, *Geology* 27, 727–730.

Further reading

Kocurek, G. (1996) Desert aeolian systems, in H.G. Reading (ed.) *Sedimentary Environments: Processes, Facies and Stratigraphy*, 125–153, Oxford: Blackwell.

SEE ALSO: adhesion; boundary layer; bounding surface; dune, aeolian; sabkha

GARY KOCUREK

INTERFLUVE

Interfluves are areas of relatively high ground that lie between adjacent valleys in a drainage basin. In its most literal sense of land 'between rivers', the term interfluves refers to undissected ridges that lie between streams. However, it is often used in a more general sense to describe generally high terrain (including lower order streams and/or VALLEYs) lying between major river systems. It has also been applied to land between glacial valleys and subglacial streams. In addition, the concept of interfluves also has application in studies of ancient sedimentary sequences, where it is used to distinguish paleo-uplands from incised valleys.

In many parts of the world the spacing between ridges and valleys is remarkably regular. For example, the coastal hills of northern California, USA, contain many examples of ridge–valley topography in which first-order DRAINAGE BASINS flanking a relatively linear main valley are quite evenly spaced (e.g. see figure 1B of Dietrich *et al.* 2003). This is also true, for example, of hollows along quartzite ridges in the ridge-and-valley province of Pennsylvania, USA. On a larger scale, many linear mountain belts show a strikingly regular space of main valleys (Hovius 1996). Neither the origins of this regularity nor the controls on valley spacing are well understood, but some clues emerge from analyses of process dynamics. Many have argued, for example, that the transition from hillslope to valley (in the absence of obvious lithologic or structural controls) reflects a switch in process dominance (Gilbert 1909; Smith and Bretherton 1972; Tarboton *et al.* 1992). Valleys, according to this argument, represent landscape zones dominated by processes that are fed by upslope surface or groundwater discharge, and whose effectiveness therefore grows with increasing drainage area (sometimes abruptly, as with Horton's (1945) concept of an OVERLAND FLOW erosion threshold). Such 'concentrative' processes can include overland flow erosion, groundwater (see GROUNDWATER) sapping, glacial scour and some forms of landsliding (see LANDSLIDE). Closer to drainage divides, processes that do not have this upstream dependence (e.g. mass movement processes such as SOIL CREEP) hold sway.

Theories based on this process-dominance concept (or the related process-threshold concept; see Kirkby 1994) successfully explain observed relationships between gradient and drainage area around valley and channel heads (Kirkby 1987; Montgomery and Dietrich 1989). They do not, however, account explicitly for the spacing between ridges and interfluves, nor do they explain the instabilities that lead to spontaneous formation of channels and valleys in the first place. The problem of spontaneous formation of channel networks was first analysed mathematically by Smith and Bretherton (1972). Using linear stability analysis, they demonstrated the conditions under which spontaneous formation of incipient channels could occur under steady, uniform sheet flow on an undissected slope. One useful result of this analysis was the notion that there will be a tendency toward CHANNELIZATION when sediment transport capacity increases more than linearly with discharge per unit contour width, so that two units of flow together carry more than twice the sediment of one. The simplifications in the Smith–Bretherton analysis, however, did not allow for prediction of the incipient spacing between channels. This shortcoming was overcome in later work (Loewenherz 1991; Izumi and Parker 2000). For example, Izumi and Parker (2000) used linear stability analysis to predict incipient channel spacing in the case of sub-critical sheet flow (see SHEET EROSION, SHEET FLOW, SHEET WASH) on a convex-upward surface of cohesive sediment. Their analysis predicted that incipient channel spacing should depend on flow depth and roughness. Using reasonable values for these parameters they were able to account for incipient channel spacing on the order of tens of metres, a value not too dissimilar from observed channel spacings of the order of 100 metres (Montgomery and Dietrich 1989; Dietrich and Dunne 1993). Numerical models of drainage basin development have also been able to reproduce hillslope–valley topography and have revealed some important controls on hillslope scale and valley density (see, for example, Willgoose *et al.* 1991; Howard 1997; Tucker and Bras 1998).

As noted above, the concept of interfluves often refers not just to individual ridges, but also to much larger areas of land lying between incised valleys. An example is the Colorado Plateau region, USA, where expanses of high-elevation, low-relief terrain are cut by abrupt, steep-sided canyons (the largest being the Grand Canyon in northern Arizona). These less-dissected upland surfaces, though still subject to active fluvial erosion, can be considered interfluves by virtue of their lesser degree of incision.

References

Dietrich, W.E. and Dunne, T. (1993) *The Channel Head, Channel Network Hydrology*, Chichester: Wiley.
Dietrich, W.E., Bellugi, D.G., Sklar, L.S., Stock, J.D., Heinsath, A.M. and Roering, J.J. (2003) Geomorphic transport laws for predicting landscape form and dynamics, in P.R. Wilcock and R.M. Iverson (eds) *Prediction in Geomorphology*, 103–132, Washington, DC: American Geophysical Union.
Gilbert, G.K. (1909) The convexity of hilltops, *Journal of Geology* 17, 344–350.
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.
Hovius, N. (1996) Regular spacing of drainage outlets from linear mountain belts, *Basin Research* 8, 29–44.
Howard, A.D. (1997) Badland morphology and evolution: interpretation using a simulation model, *Earth Surface Processes and Landforms* 22, 211–227.
Izumi, N. and Parker, G. (2000) Linear stability analysis of channel inception: downstream-driven theory, *Journal of Fluid Mechanics* 419, 239–262.
Kirkby, M.J. (1987) Modelling some influences of soil erosion, landslides and valley gradient on drainage density and hollow development, *Catena Supplement* 10, 1–14.
——(1994) Thresholds and instability in stream head hollows: a model of magnitude and frequency for wash processes, in M.J. Kirkby (ed.) *Process Models and Theoretical Geomorphology*, 295–314, New York: Wiley.
Loewenherz, D.S. (1991) Stability and the initiation of channelized surface drainage; a reassessment of the short wavelength limit, *Journal of Geophysical Research, B, Solid Earth and Planets* 96, 8,453–8,464.
Montgomery, D.R. and Dietrich, W.E. (1989) Source areas, drainage density, and channel initiation, *Water Resources Research* 25, 1,907–1,918.
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.
Tarboton, D.G., Bras, R.L. and Rodriguez, I.I. (1992) A physical basis for drainage density, *Geomorphology* 5, 59–76.
Tucker, G.E. and Bras, R.L. (1998) Hillslope processes, drainage density, and landscape morphology, *Water Resources Research* 34, 2,751–2,764.
Willgoose, G., Bras, R.L. and Rodriguez, I.I. (1991) A coupled channel network growth and hillslope evolution model; 1, Theory, *Water Resources Research* 27, 1,671–1,684.

SEE ALSO: drainage density; hillslope-channel coupling; hillslope, form; hillslope, process; valley

GREG TUCKER

INTERMONTANE BASIN

Topographic basins of various shapes and sizes are common within many uplands and low to medium-height mountains, not affected by very recent differential tectonics. They are surrounded by higher terrain on all sides and drained by rivers which typically leave the basin floor through a narrow valley. Bounding slopes are often steep and there may exist a sharp junction between the floor and the marginal slope. In the majority of cases active tectonics does not contribute to the development of the basins, hence they are features due to differential denudation. Little accumulation normally takes place within intermontane basins, except for thin alluvium along river courses and localized slope-derived deposits.

The location of intermontane basins is often influenced by lithology and structure. Basin outlines tend to follow structural lines in the bedrock, e.g. ancient faults or regional fractures, or lithological boundaries, or master joints cross each other within the basin floor. The origin of the basins can then be ascribed to selective weathering and erosion, which exploit weaker rock compartments more effectively. For example, Thorp (1967) demonstrated the existence of structure-controlled basins in granite massifs of Nigeria.

However, Bremer (1975) maintains that many basins in tropical areas cannot be explained by unequal resistance and applies the concept of divergent weathering, which is topographically rather than rock-controlled. An initial depression receives and holds more water than its drier surroundings, hence weathering mantle develops, increasing in thickness and maturity. Occasional stripping of the SAPROLITE from the adjacent slope exposes the bedrock surface, which remains dry and sheds more water into the depression. This enhances local contrast in the intensity of weathering and leads to progressive deepening of the basin and steepening of its bounding slopes, as weathering products are episodically eroded away. According to Bremer, lateral enlargement of the basins plays a minor part. Basin floors may be flat and retain a weathering mantle, or they may have some relief moulded at the WEATHERING FRONT. Low hills, groups of boulders and occasional tors may occur. Thus, basins develop through localized deep weathering and are examples of long-term, two-stage landform development.

Basins developing in the way outlined above are not to be confused with basins forming basin-and-range terrains. The latter are typical tectonic features and originate due to downfaulting of crustal blocks in areas subjected to extensional regime. Moreover, they carry a thick fill of sediment washed down from surrounding uplifted terrain.

References

Bremer, H. (1975) Intramontane Ebenen, Prozesse der Flächenbildung, *Zeitschrift für Geomorphologie Supplementband* N.F. 23, 26–48.

Thorp, M. (1967) Closed basins in Younger Granite Massifs, northern Nigeria, *Zeitschrift für Geomorphologie Supplementband* N.F. 11, 459–480.

Further reading

Johansson, M., Migoń, P. and Olvmo, M. (2001) Joint-controlled basin development in Bohus granite, SW Sweden, *Geomorphology* 40, 145–161.

Thomas, M.F. (1974) Granite landforms: a review of some recurrent problems of interpretation, in *Institute of British Geographers, Special Publication* 7, 13–37.

PIOTR MIGOŃ

INVERTED RELIEF

A type of surface morphology within which former valley floors constitute the most elevated parts of the landscape whereas pre-existing valley sides and divides have been lowered to the extent that they now form topographic lows. Relief inversion occurs when materials on the valley floors are, or become, more resistant to erosion than those underlying the adjacent slopes. This may be the case with lava flows filling concave landforms, and with DURICRUSTS developed through the induration of alluvial or lake sediments. Silcretes, calcretes and ferricretes particularly often occur in inverted position.

Typical landforms due to relief inversion are sinuous flat-topped ridges indicating former valley courses. As erosion proceeds they are reduced to isolated elongated plateaux or MESAS capped by a remnant of the more resistant material. Recognition of inverted relief may have important economic implications as former placer deposits would now occur in elevated position.

Large-scale relief inversion is often hypothesized in folded terrains (see FOLD). It is often observed that drainage lines follow axes of anticlines whereas synclines underlie ridges on both sides of a valley which seems to be in contrast to original relief. The explanation holds that once erosion reaches the softer core of an anticline, it accelerates and outpaces any lowering of adjacent synclinal structures, leading to relief inversion. Erosion hollows of the Negev Desert in Israel (*makhteshim*) are examples of this kind of inversion. A more complicated scenario invokes planation of an original mountain range in the first phase and subsequent development of inverted relief at the expense of the planation surface.

Further Reading

Goudie, A.S. (1985) Duricrusts and landforms, in K.S. Richards, R.R. Arnett and S. Ellis (eds) *Geomorphology and Soils*, 37–57, London: George Allen and Unwin.

Pain, C.F. and Ollier, C.D. (1995) Inversion of relief – a component of landscape evolution, *Geomorphology* 12, 151–165.

PIOTR MIGOŃ

ISLAND ARC

Around the western edge of the Pacific, many islands are arranged as festoons of arcs, some distance from the continent, in either single or double rows of islands. Other island arcs are the Indonesian, Caribbean, Middle America and Scotia Arcs. The arc radius ranges from about 4,000 km for Java–Sumatra to 1,500 km for Japan, and only 250 km for the hairpin Banda Arc.

The arcs face different directions: the Aleutian Arc is convex to the south, Sumatra–Java to the west, Mariana to the east, and Papua New Guinea to the north. Some 'straight arcs' such as the Solomon Islands and the Tonga–Kermadec Trench have many of the attributes of arcs without the curvature.

Some arcs are separated from the continent by backarc basins (Japan). The Scotia Arc has no backing continent, only the Pacific seafloor. The Caribbean Arc is backed by the straight Middle America Arc. The Mariana Arc is separated from the Philippine Arcs by a complex of backarc basins.

The standard arrangement of landforms in a simple arc is: continent; backarc basin; arc – forearc basin; accretionary prism; trench; ocean (Figure 94). The simplest arcs are lines of active volcanoes (e.g. Kurile Arc), which may be subdivided on the basis of underlying rocks into volcanic or continental based. The commonest type of double arcs comprises two rows of volcanoes, one older than the other. The older, usually outer line may be wholly or partially covered by limestone. The Sumatra–Java Arc is volcanic, with an outer arc of sedimentary rocks.

Topographically, the front of an island arc is generally rising while the back of the arc is sinking. In New Britain the front of the arc is rising, with flights of uplifted coral terraces, while the other side has drowned coasts, indicating sinking. Many arcs have large, mountainous islands built of sedimentary, metamorphic and granitic rocks (Japan, Papua New Guinea). They were eroded to a planation surface before the modern mountains were uplifted. The Virgin Islands display an erosion surface at about 300 m, and many terraces. Uplift occurred within the past 2 million years.

On the ocean side of arcs lie deep trenches. The greatest known depth in the ocean is the Mariana Deep (11,035 m). Traverses of trenches reveal numerous normal faults and graben. Trenches contain variable amounts of sediment: part of the Chile Trench is virtually empty; the Aleutian Trench contains up to 4 km of undeformed horizontal layers of sediment. An area of relatively shallow water separates the islands of the arc from the trench bottom. This is called the arc–trench gap. It is usually over 100 km wide, and is 570 km in the eastern Aleutian Arc. It is underlain by thick, generally horizontal sediments of the forearc basin.

At the outer edge of the forearc basin there is a marked break in slope, and the steep slope bounding the trench may constitute an accretionary wedge. The accretionary wedge (or prism) is a package of highly deformed sediment, and perhaps oceanic basalt, presumed to be scraped off the downgoing slab and accumulated at the edge of the overriding slab. Alternatively, the deformation structures can be interpreted as gravity-tectonic structures with décollement (also called detachment faulting) and thrusting. The whole area from the volcanic arc to the ocean, including the trench, is called the forearc. An alternative plate tectonic scenario is that the sediment, instead of being scraped off, is subducted beneath the arc.

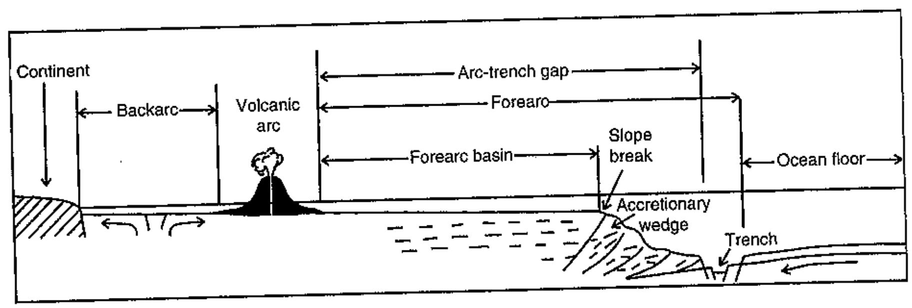

Figure 94 Nomenclature of features in a typical island arc

Behind island arcs are shallow seas known as backarc basins. Backarc basins are spreading sites (Taylor 1995). The Japan backarc basin has up to 2 km of sediment. The curved shape of the arc is not easily related to the angular pattern of multiple spreading sites behind the arc (e.g. Scotia Arc). Nor is the curve related to the approaching Pacific Plate: part of the Aleutian Arc is parallel to the movement, and part is almost perpendicular.

Active volcanoes on arcs frequently erupt with great violence, as at Krakatau (1883), Mont Pelée (1902) and Mount Pinatubo (1991). Stratovolcanoes are the commonest type, with numerous calderas. Andesitic volcanoes predominate in arcs, but some notable volcanoes are basaltic, including Mt. Fuji, Japan. Basalt makes up 70 per cent of the South Sandwich Islands. Arcs with volcanoes were formed in the past, as they are today, sometimes making double arcs. In Papua New Guinea the New Britain Arc has Quaternary volcanoes on the inside and the Palaeogene arc on the outside, but the New Ireland Arc has inner Palaeogene volcanoes and an outer arc of Quaternary volcanoes.

In some arcs (New Hebrides, Middle America) normal faults break the Earth's surface into fault blocks. Normal faults are also common in the trenches and backarc basins. Strike-slip faulting is important in some arcs, with lateral displacements of many kilometres along faults roughly parallel to the arc. These faults are a major feature in the geomorphology of Sumatra and Java. The Middle America Arc is offset in Nicaragua by a strike-slip fault oblique to the arc, which is a continuation of the Clipperton Fracture of the Pacific floor.

The American arcs are somewhat different from the rest. The Caribbean Arc lies east of the Caribbean Plate. This is bounded to the north by what looks like streaked-out bits of North America, consisting of continental rocks and making the islands of the Greater Antilles as far as Puerto Rico. The southern side is similarly made of rocks like those of South America. The true arc is the north–south part of the Leeward Islands, a typical double arc with the Limestone Caribees to the east and Volcanic Caribees to the west. The western side of the Caribbean Plate is bounded by the Middle America Arc, with the Middle America Trench on the Pacific side. The trench has no accretionary prism, and sediments are horizontal with no evidence of compression. The arc has some old continental rocks, and on the western side old volcanic rocks. Topographically the arc is a plateau tilted up to the west, but there is a down-faulted strip on the Pacific side, and this is where the current volcanic arc lies. The Isthmus of Panama is not part of the volcanic arcs, but consists of block-faulted continental rocks.

The Scotia Arc lies east of the Scotia Plate, which has continental rocks from South America streaked out on the northern side, and Antarctic rocks on the southern side. The South Sandwich islands make a true arc, with a spreading site behind that separates the Scotia Plate from a much smaller Sandwich Plate which is less than 8 million years old. The islands are volcanic, rugged and glaciated. The Scotia Plate is separated from the Pacific by a transform fault.

A negative gravity anomaly lies over trenches, and a positive anomaly on the continental side about 115 km distant. Arcs are characterized by earthquakes. The foci sometimes appear to fall in a zone (called the Benioff zone), about 50 km thick and reaching depths of several hundred kilometres. This dips towards the concave side at varied angles, and is vertical under the Mariana Arc. It is presumed to mark a slab being subducted at the trench. In some arcs (e.g. Aleutians) the Benioff zone emerges not at the trench, but on the arc–trench gap. The relation between arc, trench, gravity anomalies, earthquakes and volcanoes is in fact quite variable.

Some arcs run aground onto continents. The Sumatra Arc goes via the Andaman Islands to Burma (Myanmar), and continues as the landbound arc of the Himalayas of about the same size. The Aleutian Arc continues in Alaska; and the southern arm of the Caribbean Arc in Colombia. This is significant in theories that relate island arcs to mountain building, a major concept in PLATE TECTONICS. Since backarc spreading increases the space between continent and arc it might seem difficult to make the two collide, but Hamilton (1988) wrote that island arcs are 'conveyor-belted' towards subduction zones, so that island arcs collide with one another and with continents. Island arc concepts are often used to interpret mainland features including ancient structures and modern mountains such as the Apennines.

The ruling theory of plate tectonics explains island arcs as places where oceanic crust is subducted at trenches. This produces the paradox of having a compressive mechanism, when the normal faults in trenches, islands and backarc basins all indicate tension. Other suggested mechanisms of arc formation include geotumours, mantle diapirism, subsided blocks at the trench, and surge tectonics.

References

Hamilton, W.B. (1988) Plate tectonics and island arcs, *Geological Society of America Bulletin* 100, 1,503–1,527.

Taylor, B. (ed.) (1995) *Backarc Basins, Tectonics and Magmatism*, New York: Plenum.

Further reading

Wezel, F.-C. (1986) *The Origin of Arcs*, Amsterdam: Elsevier.

The Island Arc (Blackwell) is devoted to arcs and related topics and appears four times a year.

SEE ALSO: plate tectonics

CLIFF OLLIER

ISOSTASY

A term introduced in 1882 and derived from the Greek words *iso* and *stasis* that means equal standing. It is used to describe a condition to which the Earth's crust and mantle tend, in the absence of disturbing forces – a condition of rest and quiet in which the lighter crust floats on the dense underlying mantle (Watts 2001). However, isostasy is disturbed by such processes as sedimentation, erosion, volcanism and the waxing and waning of ice sheets (see GLACIAL ISOSTASY). In its simplest form the isostasy concept envisages that rigid blocks of crust, buoyantly supported by the underlying fluid medium of the mantle, are free to move vertically until their weight is balanced by their buoyancy ('isostastic equlibrium'). Differences in the density and thickness of the crust are to a considerable extent responsible for variations in the isostatic adjustment of the lithosphere. If it is in isostatic equilibrium, one portion of the lithosphere will stand higher than another, either because of a lower density (the so-called Pratt model) or because it is of the same density but thicker (the Airy model), or from a combination of both.

Isostatic equilibrium is disturbed by various geomorphological processes. Erosion makes crustal blocks thinner and lighter in weight so that an eroding mountain will tend to rise to maintain equilibrium. Conversely, deposition of sediments, as for example in a delta, represents an added load, so that sinking tends to occur. Extension of the crust by rifting thins it whereas compression in mountain-building thickens it. Loading and unloading by waxing and waning glaciers also affects isostatic equilibrium as do changes in the volume of water in the ocean basins or in lakes – hydro-isostasy (Bloom 1967). Similarly, the extrusion of large amounts of volcanic material can weigh down the crust and cause subsidence to occur.

Isostasy is central to understanding a range of geomorphological phenomena. For example, areas that were once loaded by great ice sheets (e.g. Fennoscandia) are now areas of uplift as they recover from the load of the ice. Conversely areas peripheral to the ice sheets, which bulged up during glacials, are now areas of subsidence (e.g. the southern North Sea area). Likewise, the development of GUYOTS, seamounts and atolls in the Pacific may be related to subsidence caused by volcanic eruptions loading the crust (McNutt and Menard 1978). Load-induced vertical movements have a profound effect on deltas. In particular, subsidence increases the overall water depth and so increases the accommodation space that is available for the prograding sediments. The relief characteristics of rift valleys, features caused by crustal extension, also show the effects of isostasy through their association with broad topographic swells. The rift flanks of passive margins show evidence of uplifted flanks and erosional unloading (as in Namibia) and the presence of rift flank uplifts may explain the deflection of drainage systems towards the continental interiors (e.g. the Kalahari) (Gilchrist and Summerfield 1990). The shorelines of great palaeo-lakes, like Lake Bonneville in the southwest USA, have deformed as the crust has adjusted to the removal of the weight of the lake as it desiccated (Crittenden 1967).

The continental shelves were depressed as the weight of water from rising postglacial sea levels (the Flandrian Transgression) was applied to them. Finally, erosion that deepens and widens river valleys but does not erode the peaks to the same degree, will reduce the mass of an area and so drive isostatic uplift. As a consequence, the altitude of the peaks could increase at the same time that the mean height of a region is decreasing (Molnar and England 1990).

Illustrations

Figures

Plates

Tables

J

JOINTING

Joints are cracks in rocks formed by stress that results from tectonic events, cooling, or isostatic rebound. They range in length from millimetres to kilometres. In outcrop, joints can be small hairline cracks only millimetres in length or long open fissures a metre or more across. They may be open or filled; common fillings are soil and clays. Joints can also be sealed as a result of hydrothermal activity. They are distinguished from faults (see FAULT AND FAULT SCARP) by the lack of movement between the two sides of the joint. Most joints are tension fractures.

Joints occur in sets, groups of nearly parallel fractures that formed under the same stress regime. At least three joint sets commonly occur in most outcrops: in sedimentary and schistose/foliated metamorphic rocks. There is typically one set parallel to layering and two mutually perpendicular sets normal to layering. In massive rocks like granite, one set is usually horizontal and the others are vertical or steeply dipping. Columnar joints in basalts (e.g. the Giants Causeway, Northern Ireland) are a special example of cooling joints. These joints bound areas where cooling fronts meet so that a hexagonal crack pattern forms where the boundaries converge.

Joints exert significant control on landform shape. Most granite landforms, such as domes (e.g. Stone Mountain, Georgia, USA and Ayers Rock, Australia) and TORS (e.g. Haytor Rocks, Dartmoor, UK), are joint controlled. Other joint-controlled landforms not dependent upon lithology include geos and gulls. A geo is a deep, narrow cleft or ravine (see GORGE AND RAVINE) along a rocky sea coast that is flooded by the sea. Valley orientation in general often parallels major joints. A gull is a joint that opens on escarpments because of tension produced by cambering (see CAMBERING AND VALLEY BULGING). Grikes in LIMESTONE PAVEMENTS (e.g. the Burren, Co. Clare, Eire) are joints enlarged by solution that separate clints, the raised portions of these pavements.

Joints are also very important with respect to both MECHANICAL WEATHERING and CHEMICAL WEATHERING, and slope stability. Joints are the most important zones of weakness in any given rock mass, and provide primary access for moisture to enter rock. A dense pattern of closely spaced joints will thus hasten chemical weathering, leaving upstanding areas where joints are more widely spaced. The WEATHERING FRONT often occurs along a horizontal joint. In addition, most limestone CAVEs originate as joints, and sink holes commonly occur at joint intersections. Corestones are remnants of joint blocks, often occurring in a matrix of weathered debris or as boulder fields after the weathered matrix has been stripped. Furthermore, the stability of a given slope can often depend on the orientation of the joints. Sliding is likely to occur if the joints dip toward the slope face, for instance, and toppling occurs along vertical joints oriented parallel to the slope. Joints also commonly form the slip surface in larger rotational landslides. DRAINAGE PATTERNs are also often controlled by the joint pattern of the underlying rocks, e.g. rectangular drainage patterns in limestone, and drainage density may reflect joint density.

Further reading

Engelder, T. (1987) Joints and shear fractures in rock, in B.K. Atkinson (ed.) *Fracture Mechanics of Rocks*, 27–69, London: Academic Press.

Gerrard, A.J.W. (1974) The geomorphological importance of jointing in the Dartmoor granite, in E.H. Brown and R.S. Waters (eds) *Progress in Geomorphology*, Institute of British Geographers Special Publication 7, 39–50.
——(1988) *Rocks and Landforms*, London: Unwin Hyman.
Hencher, S.R. (1987) The implications of joints and structures for slope stability, in M.G. Anderson and K.S. Richards (eds) *Slope Stability*, 145–186, Chichester: Wiley.
Twidale, C.R. (1982) *Granite Landforms*, Amsterdam: Elsevier.

SEE ALSO: bedrock channel; bornhardt; granite geomorphology; inselberg

JUDY EHLEN

K

KAME

The term kame is derived from the Scottish word *kaim* and was introduced by Jamieson (1874) to mean a short, steep-sided mound or ridge of water laid sands and gravels deposited from melting ice. As a single term it cannot be applied to a specific landform, depositional process or sediment type. A variety of landforms resulting primarily from glacifluvial deposition in association with melting and buried ice are now recognized as kame forms. They include kames, kame terraces, kame complexes, and kame and kettle topography. These landforms occur in subglacial and ice-marginal environments (Holmes 1947; Price 1973; Bennett and Glasser 1996; Benn and Evans 1998).

Kames form subglacially or ice-marginally as small hills or ridges that may occur as isolated features or in a group of similar forms. Some form through deposition of sand and gravel at the base of a glacier by meltwater descending from the ice surface in a moulin, or from an englacial water body. Stream deposition within a channel bounded by ice walls may produce a short kame ridge. Other kames form as small delta-kame surfaces at the ice margin or terminus, and frequently contain beds of sand and gravel, debris flow diamicton and laminated silts and clays. Melting of adjacent and underlying ice results in normal faulting of the sediments on mound margins and sagging and folding of beds. Clay and silt beds often have slump, flow and load structures.

Kame terraces are usually formed either at the margin of the glacier between the debris-mantled decaying ice and the valley slope, at the terminus of the glacier along the ice front or between the decaying ice and an obstructing moraine. A marginal kame terrace surface may slope slightly down-glacier or towards the valley slope. The primary terrace may be fragmented into several sections due to collapse of underlying ice. Several kame terraces can be developed on a slope as the glacier margin recedes. The depression formed between the ice margin and valley slope diverts the meltwater laterally along the ice margin. The depression may be temporarily filled with water to form a narrow marginal lake. Glacifluvial sediment transported along the ice margin and from the ice edge is deposited either onto the buried ice on the floor of the depression or into the lake, or both.

The sediments in the kame terrace vary rapidly in texture and structure laterally and vertically. They consist predominantly of horizontal to gently inclined alternating beds of sand and small gravel due to glacifluvial deposition, but may exhibit laminated silts and deltaic foreset beds due to deposition in relatively deep water. On the ice proximal side of the terrace interbedded masses of debris flow diamicton (flow till) from the ice margin may occur while on the valley side paraglacial deposition of alluvial fans and debris flows can occur due to high discharges and snowmelt in adjacent valleys and on slopes. Melting of buried ice during and subsequent to sediment deposition results in formation of kettle holes that are more frequent near the former glacier margin and sometimes contain small lakes. Removal of support from the sediment by melting of marginal and buried ice results in collapse of the terrace edges to give a crenulated form, due to debris flows and landslides in the ice contact zone. Normal fault, fold, slump and flow structures may be formed within the sediments.

Kame terraces of the terminal zone usually form steep-sided ridges or steeply bounded low plateau surfaces. They may form where ribbon lakes are developed along the ice edge and are filled primarily with glacifluvial sands and gravels. They can also form where a small lake is impounded between the decaying ice and a bounding moraine. A small stream may flow from the ice bringing sand and gravel but much of the sediment consists of laminated silts interbedded with flow diamictons from the ice surface. Melting of underlying and removal of supporting marginal ice usually results in the sediments exhibiting fault, fold, slump, flow and load structures.

Kame complexes occur as a number of steep-sided mounds or ridges usually within a limited area. They result from the deposition of sands and gravels by meltwaters descending from within the ice to the glacier bed where the deposits accumulate in numerous cavities. Since deposition of kame complex sediments mainly takes place subglacially, the glacifluvial cores of the ridges are often discontinuously draped by diamicton resulting from meltout and flowage from the overlying ice during its final decay. The sediments may show faulting, folding, sag, slump and flowage structures.

Kame and kettle topography is recognized in a landscape by the occurrence of many steep-sided hills or short ridges closely juxtaposed and separated by relatively deep circular, oval or elongate depressions. Several depressions may contain ponds. Kame and kettle topography differs in origin from kame complexes by the glacifluvial sedimentation having occurred over abundant buried ice at and beyond the glacier margin. Ten to over twenty metres of glacifluvial sands and gravels are often deposited and the dips and directions of the beds are highly variable being related to deposition into depressions in locally stagnant ice. Melting of the buried ice causes normal faulting and folding of the sediments and a characteristic inversion of the topography with the kames occupying the former sites of deposition and the kettle depressions and lakes the sites of the thickest buried ice masses.

References

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.

Holmes, C.D. (1947) Kames, *American Journal of Science* 245, 240–249.

Jamieson, T.F. (1874) On the last stage of the Glacial Period in North Britain, *Quarterly Journal of the Geological Society of London* 30, 317–338.

Price, R.J. (1973) *Glacial and Fluvioglacial Landforms*, Harlow: Longman.

ERIC A. COLHOUN

KAOLINIZATION

Kaolin ($Si_2Al_2O_5(OH)_4$) is a clay mineral which can be produced by prolonged weathering during which chemical elements are progressively lost. The more complex clay montmorillonite ($(Al_{1.7}Mg_{0.3}Si_{3.9}Al_{0.1}O_{10}(OH)_2)$) is comprised of not only the basic crystal lattice structure of aluminium and silicon atoms but it also contains magnesium. Kaolin is simply comprised of the more resistant silicon and aluminium, the more soluble elements such as magnesium having been lost through weathering. Thus kaolinization represents a process of the loss of the more soluble chemical elements producing a much simpler crystalline clay over time and can be produced by the prolonged weathering of feldspars, micas and other primary aluminosilicate minerals which may originally have contained magnesium, calcium, sodium, iron and potassium. Kaolinization thus refers to the changes from primary rock aluminosilicate minerals through weathering and their transformation to kaolin as a residual clay mineral.

Kaolin is formed under conditions of slight acidity and free drainage which remove most of these other cations. Free drainage and high rainfall are important factors in the removal of the soluble cations and in areas with annual rainfall of below around 500 mm montmorillonite clays dominate, between 500 and 1,500 mm kaolin clays dominate, with iron oxides also occurring at rainfalls above 1,500 mm (Thomas 1974). The SOLUBILITY of kaolinite is least at pH 6 but increases both as pH rises and falls from this value; so as pH departs from 6 and under strongly oxidizing conditions in the humid tropics, intense weathering tends to remove most of the silica, leaving the oxides of iron or aluminium as residual minerals rather than kaolin.

The distribution of kaolinized mineral material through a vertical weathering profile characteristically shows an increase towards the surface with a concomitant decrease in primary minerals. In a study of weathering profiles in Ghana (Ruddock 1967), some 60 per cent of particles finer than 40 μm were kaolinite at a depth of some 5–6 m and this dropped to around 30 per cent at around 30 m depth while feldspar constituted 25 per cent of the particles over 40 μm in size at 30 m and was absent at 5–6 m. The zone of greatest kaolinization may be around the zone of the water table, with kaolin often accumulating at the base of the vadose zone. The presence of clays at this depth may reduce vertical permeability to water and thus facilitate lateral water movement.

It has been suggested that the presence of kaolin is indicative of past tropical weathering conditions. What is more the case is that kaolin is indicative of prolonged weathering with slightly acid conditions, high rainfall and good drainage to facilitate cation loss. While it is thus true that kaolin is present in currently tropical areas, the presence of kaolin is only indicative of prolonged weathering under appropriate and stable conditions. It should be remembered that kaolin is also produced by metamorphosis in proximity to molten magmas.

References

Ruddock, E.C. (1967) Residual soils of the Kumasi district in Ghana, *Geotechnique* 17, 359–377.

Thomas, M.F. (1974) *Tropical Geomorphology*, London: Macmillan.

STEVE TRUDGILL

KARREN

Karren are surface solutional weathering features of varied size and shape, found on karstifiable (see KARST) rocks, usually limestones, dolomites and dolomitic limestones, but including gypsum, rock salt and silicates. The French word is *lapié* but the German *karren* generates terms now widely accepted. Wherever karstifiable rocks exist there will generally be karren features. Early research, in the nineteenth century, focused on observations in Alpine Europe. Later knowledge of karren extended to many climatic zones. Considerable experimental and field research to examine form and conditions of formation has been carried out.

Karren extend from the nanoscale of individual features to landform complexes measured in metres or tens of metres, the latter often having smaller karren on their surfaces. This rock sculpturing by DISSOLUTION, involves complex, and, sometimes, complete surface patterning. Removal (negative) forms are fundamental, e.g. pits, rills, or runnels, created directly by process (Ford and Lundberg 1987). Other karren are remnant positive forms, e.g. flachkarren (clints) and spitzkarren. Microscale forms, e.g. pits and rillenkarren, are on a centimetre scale or less. Large solution runnels, carved pinnacles, karrenfelder, giant grikelands, LIMESTONE PAVEMENTS and other complex polygenetic assemblages lie at the other extreme. Features vary in basic shape, e.g. some are linear, others circular. The largest linear solution forms can be tens of metres long although the commonest are tens of centimetres long. Width of features is also usually in tens of centimetres. Restrictions on dimensional development are often due to downslope available rock distance. Depths likewise vary considerably due partly to topographic opportunity. Small-scale surface roughness varies, with some karren characteristically smooth, others spikey or rough (Plate 66). Karren-like forms are also found underground.

Karren development is principally affected by: the type and intensity of chemical processes attacking the rock; rock characteristics including intrinsic lithology, disposition (structure) and relation to surrounding topography; the nature of any cover material; and the time available, and the changes in conditions during that time, for processes to act.

Karren classification has been attempted. Some schemes examine processes: Bögli's (1960, trans. in Sparks 1971) involves the effects of the main carbonate solution processes, depending on whether the limestone is exposed to the air, partly

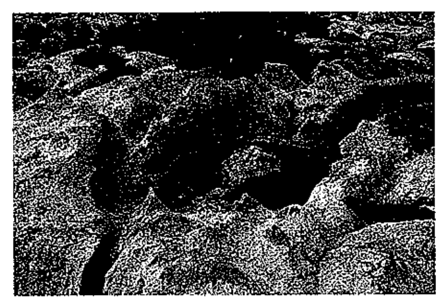

Plate 66 Spitzkarren in the Triglav area of the Julian Alps, Slovenia

covered, or entirely covered by soil. This cover factor fundamentally influences solutional processes on limestone, affecting the amount of CO_2 available for solution and the time period and speed of solution. Many karren are examples of BIOKARST as they are fundamentally the result of biological corrosion.

However, many karren are not simple genetically, for they reflect past as well as present processes, and many forms are polygenetic. Thus Ford and Williams's 1989 classification is subdivided to allow for genetic factors, rather than a purely genetic system. They retain descriptive karren terminology, finding form intuitively useful. They distinguish circular karren forms, those which are linear and controlled by structure, linear forms controlled hydrodynamically, and polygenetic forms. However, these distinctions may be blurred; for example, after karren initiate hydrodynamically, field evidence suggests that small-scale lithological structures affect development. Weaknesses affect fluid movement, and may interrupt karren, causing capture or change. The importance of lithology at the microscopic scale has been stressed by researchers such as Goudie *et al.* (1989). Slope significantly influences karren, especially the pattern and complexity of branching networks.

It is useful to describe the main forms. *Rillenkarren* are tiny gravitomorphic packed channels, starting from a crest and extinguishing downslope. Width is about 1 to 3 cm, and length a few tens of cm. They are separated by sharp ridges.

Rinnenkarren and *rundkarren* are Hortonian channels with an unrunnelled catchment surface above their commencement points. They enlarge downslope depending on available rock. Width varies but rundkarren often stabilize at 20 to 30 cm and rinnenkarren are narrower. High in the long section, rundkarren develop parabolic cross sections, deepening downstream before stabilizing at the bottom end. Occasionally overdeepening develops where the runnel flows into a grike. Rundkarren are smooth features with well-rounded crests, whereas rinnenkarren are distinguished by sharper crests. The former result from covered conditions and soil removal makes them visible. Rinnenkarren generate much debate, but are considered to be formed in free or, possibly, half-free conditions. Flow forms may meander (*meanderkarren*) depending partly on slope. More complex branching of rundkarren is found on gentler slopes.

Trittkarren are step or heel-print-like features, about 10 to 30 cm in scale, essentially features of very gentle slopes. Their headwall is arc-shaped, they are flat floored and they open downslope (Vincent 1983). Trittkarren appear related to the ripples forming from sheet flow across a gentle sloping rock surface (White 1988).

Spitzkarren are peak-shaped features remaining from surface solution widespread over horizontal or gently sloping surfaces. Their sides are carved by rillenkarren. Their diameter is typically 50 cm and their height about 10 cm. They tend to merge into other positive features, pinnacles in particular, and are essentially polygenetic (Plate 66).

Kluftkarren are clefts, fissures or grikes. These are the major splits into limestone surfaces formed by widening, deepening and eventually merging of small solution features developing along linear weakness in the rock. Mature examples may run considerable lengths, e.g. several metres. Depth varies with bed thickness. By definition grikes should have broken through at least the top bed of an outcrop. This distinguishes them from runnels, which may develop into grikes (Plate 67, and see LIMESTONE PAVEMENT, Plates 73 and 74).

Kamenitzas are solution basins or pans (see WEATHERING PIT). Size varies enormously, but small ones have flat floors and scalloped edges about 3 cm deep, with diameters of several cm or more. Many are tens of cm in diameter with the largest examples measurable in metres. Development is, however, limited, as when the pan becomes drained either at the surface over

Plate 67 Sloping limestone pavement (12° to 15°) on Farleton Fell, Cumbria, UK: showing strong karren formation on a wide range of clints across several limestone beds

a rim, or from beneath via solutionally opened cracks in the floor. The feature may become more complex; perhaps part of a staircased runnel, or of a complex hole into which several runnels drain. Kamenitzas occur in both free and covered conditions. True kamenitzas enlarge outwards at their rims where solution conditions remain ideal. Kamenitzas may merge into kluftkarren, or coalesce, leaving peaky sharp Spitzkarren.

Kluftkarren also merge into larger weakness-oriented features, e.g. bogaz, lapiés wells, etc. Several joints crossing focuses processes, resulting in mill-like features possibly several metres across and complex in form. This is demonstrated in limestone pavement areas in the UK, and in high mountain areas such as north Norway and the Alps, where large amounts of glacial meltwater may have enhanced them.

Terminology cannot encompass the full variety of natural sculptured forms. Features grade into each other, cut across, develop idiosyncratically according to local conditions and deteriorate when ideal formation conditions change. Destruction of rundkarren, formed under soil, illustrates this: on soil removal their smooth surface undergoes etching in subaerial conditions into sharper rougher features, e.g. kamenitzas or rillenkarren. Destruction may include mechanical weathering effects involving freeze and thaw, for example, and ends with broken rubble fields over underlying intact rock layers. However, another change in conditions could find water flow, or even a return to soil cover: both would alter the karren again. In karst areas such changes of local conditions can happen especially easily and quickly because of capture of active-process locations.

Rock type is fundamental to karren development. Strong, pure limestones or dolomites produce the best development. On rock salt, karren form easily due to high solubility but only persist in relative aridity. Gypsum karren are intermediate in persistence as gypsum's solubility is such that karren develop in humid conditions, but these conditions also favour their destruction. Silicate rocks are only slightly soluble; only in prolonged warm and wet conditions will karren, like other karst features, form. Sandstones in temperate areas may display weak karren but development is only significant in the wet tropics. Research has placed a general timescale for development of limestone karren in thousands of years, gypsum features in hundreds of years and rock salt karren in tens of years (Mottershead and Lucas 2001).

Other lithological factors influence particular karren forms. Research has considered karren initiation at tiny weaknesses or variations in the rock (Moses and Viles, in Fornos and Gines 1996). Chance unevenness allows rainwater to pond and start surface solution, simple plants may then develop. If soil develops so do higher plants and accelerated biological corrosion becomes very important, especially in warm, wet locations. The forms produced can be very striking. On bare rock, flow processes down any slope, however smooth, result eventually in flow-concentration into channels. A slight slope gives flow rather than ponding, a slight dip can give ponding before flow.

Sense can be made of karren by considering their form, mode of origin and development conditions. However, field situations demonstrate that, although 'perfect' examples of types are found, there will always be a wide spectrum, and merging of features is both possible and common even without changes in external factors such as climate.

References

Ford, D.C. and Lundberg, J. (1987) A review of dissolution rills in limestone and other soluble rocks, *Catena Supplement* 8, 119–140.

Ford, D.C. and Williams, P.W. (1989) *Karst Geomorphology and Hydrology*, London: Unwin Hyman.

Fornos, J.J. and Gines, A. (eds) (1996) *Karren Landforms*, Palma: Universitat de les Illes Balears.

Goudie, A.S., Bull, P.A. and Magee, A.W. (1989) Lithological control of rillenkarren development in the Napier Range, Western Australia, *Zeitschrift für Geomorphologie N.F. Supplementband* 75, 95–114.

Mottershead, D. and Lucas, G. (2001) Field testing of Glew and Ford's model of solution flute evolution, *Earth Surface Processes and Landforms* 26, 839–846.

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

Vincent, P.J. (1983) The morphology and morphometry of some arctic Trittkarren. *Zeitschrift für Geomorphologie N.F. Supplementband* 27(2) 205–222.

White, W.B. (1988) *Geomorphology and Hydrology of Karst Terrains*, New York: Oxford University Press.

SEE ALSO: biokarst; dissolution; karst; limestone pavement.

HELEN S. GOLDIE

KARST

Karst is terrain with distinctive hydrology and landforms arising from a combination of high rock solubility and well-developed secondary porosity (Ford and Williams 1989). It is commonly associated with carbonate rocks such as limestone, marble and dolomite and is well known for features such as caves, enclosed depressions, fluted rock outcrops, underground rivers and large springs. It takes its name from a limestone region in the northern Adriatic inland of Trieste on the Slovene–Italian border where such features are particularly well developed. The region is called *kras* (Slovenian) or *carso* (Italian), but this was Germanicized to *Karst* in the period of the Austro-Hungarian Empire, when the first scientific studies were made of the region's geomorphology and hydrology. By extension, other areas with similar features are also referred to as karst, and this includes places where it is developed on other soluble rocks such as gypsum and rock salt (see GYPSUM KARST; SALT (EVAPORITE) KARST). This discussion focuses on karst in carbonate rocks. These outcrop over about 12 per cent of the ice-free continental areas (Figure 95), with well-developed karst covering about 7–10 per cent of the continental area. Karst also develops beneath the surface when karst rocks are interbedded with other lithologies; this is known as *interstratal karst*. Limestone (as opposed to karst) geomorphology refers to landscapes developed on carbonate rocks, and includes landforms that are not necessarily produced by karst processes (e.g. coral atolls, glacial troughs).

Cvijić (1960) defined different morphological types of karst, including *holokarst* where karst reaches its fullest development in thick carbonate rocks that extend below sea level, for example in the extensive limestones of the Dinaric region; *merokarst* where karst development is evident but rather poorly expressed, by virtue of not very suitable lithologies or rather thin limestones, such as in the chalk of northern France or in some areas of Jurassic limestones such as the Swabian Alb of Bavaria or the Cotswold Hills of Britain; and transitional types where the carbonate rocks are quite thick and well karstified, but underlain by or interbedded with non-carbonate formations, as in the Causses of France. However, these terms are seldom used. When the imprint of now dry river valleys is evident in the landscape, it is sometimes referred to as *fluviokarst*.

The occurrence of pure carbonate rocks with high solubility is insufficient to produce karst, because the structure, density and thickness of the rock are also important. The carbonate rocks in which karst reaches its best expression are thick, dense, pure (> 90 per cent calcium carbonate) and massive with low primary porosity, but with well-developed secondary porosity along fissures such as joints, faults and bedding planes. Even pure soluble rocks such as coral and chalk have relatively poorly developed karst, because their very high primary porosity (which can be 30–50 per cent) leads to diffuse groundwater flow, which is not conducive to extensive cave and closed depression development, though they have some. Impure carbonate rocks such as argillaceous limestones hardly support karst, partly because they tend to have high primary porosity but especially because insoluble residues inhibit the growth of secondary porosity by clogging groundwater pathways as they form. Nevertheless, there is a spectrum of rock types and degrees of purity, with a corresponding spectrum of karst development.

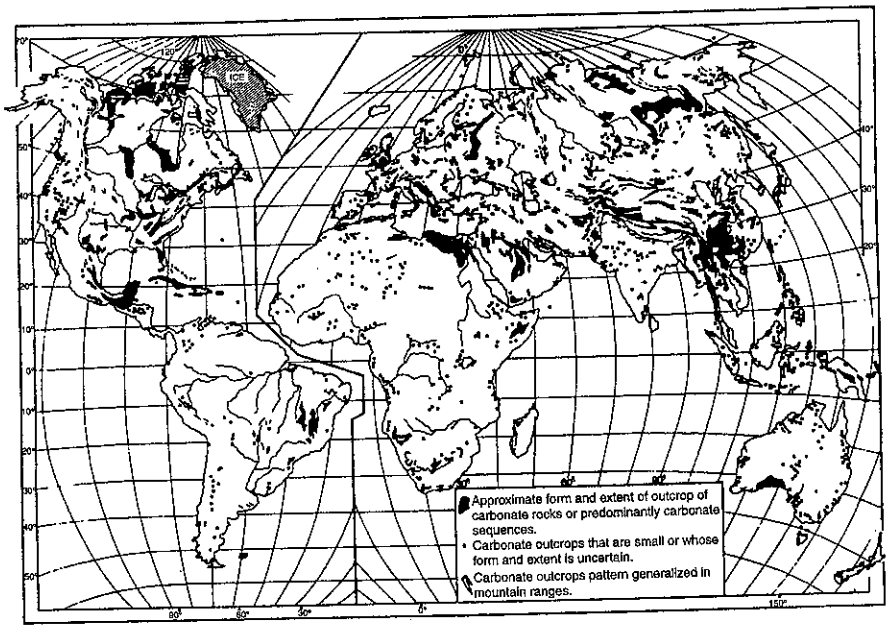

Figure 95 Global distribution of carbonate rock outcrops. Karst occurs over most of this area and also in subcrop beneath various coverbed lithologies (from Ford and Williams 1989.)

Chemical processes

The dominant process that produces karst features is the solution of the rock by rainwater. The chemical process is DISSOLUTION (or corrosion); in a carbonate karst context the process can be summarized as:

$$CaCO_3 + H_2O + CO_2 = Ca(HCO_3)_2$$

Calcium carbonate, $CaCO_3$, dissolves in the presence of water and carbon dioxide ($H_2O + CO_2 = H_2CO_3$ or carbonic acid) to yield the more soluble calcium bicarbonate, $Ca(HCO_3)_2$, which is readily transported away in solution.

The process of carbonate rock solution can be conceptualized as operating in a situation in which two hydrological and geochemical subsystems interact. The hydrological cycle provides the main source of natural energy that powers the system and drives the evolution of karst, because water is the solvent that dissolves karst rock and then carries it away in solution. Geochemical processes control the rate of dissolution (the speed with which solid rock is converted into ions in solution), which in a carbonate karst depends very strongly on strength of acidification by dissolved carbon dioxide during its passage through the atmosphere and soil layer before making contact with the limestone. The concentration of CO_2 in the open atmosphere is about 0.03 per cent by volume, whereas it is commonly 2 per cent in the soil and can even reach 10 per cent. A factor of 100 in the concentration of CO_2 results in ~5 times increase in the solutional denudation rate (White 1984). Although this is important, the amount of rainfall is even more significant, the wettest places in the world having the fastest rate of limestone solution. For example, limestone denudation by solution processes has been estimated as high as $760\,m^3\,a^{-1}\,km^{-2}$ (cubic metres per year per square kilometre of limestone outcrop) in very wet places such as parts of Papua New Guinea where rainfall can reach 12,000 mm per year, but as low as $5\,m^3\,a^{-1}\,km^{-2}$ in some arid zones like the Nullarbor Plain in southern Australia with rainfall of less than 350 mm per year. The amount of solutional attack on the limestone rock therefore depends on the concentration of the solute (determined by biogeochemical processes) and the volume of solvent (determined by the rainfall). The solute load of a karst spring is the product of its discharge and the concentration of limestone salts in solution.

Biochemical and physical processes associated with various organisms also assist in weathering limestones, especially in the intertidal zone, and produce a suite of landforms known as BIOKARST.

Landscape development

A conceptual model of the karst system is presented in Figure 96. Karst evolution is explained by White (1988), Ford and Williams (1989) and Gabrovšek (2002), and is summarized in Williams (2003). In order for major karst landforms such as enclosed depressions and caves to develop, the rock removed in solution must be carried right through the body of karst rock and be discharged at springs. Thus the development of an underground plumbing system is a necessary precursor to surface landform evolution. When a continuous conduit of 5–15 mm extends right through the rock, the drainage can become turbulent and

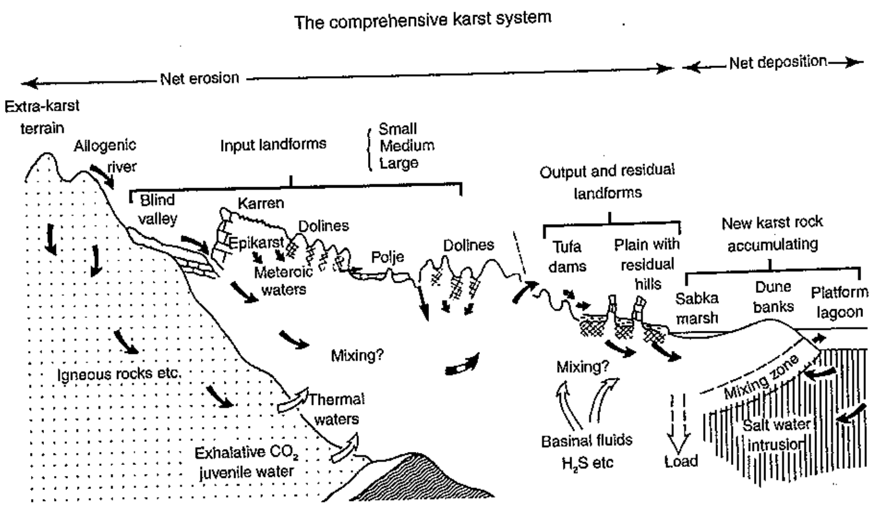

Figure 96 A conceptual model of the comprehensive karst system (from Ford and Williams 1989.)

then can begin to discharge fine insoluble particles. It takes the order of 5,000 years for a conduit of 1 km length to develop up to this size (White 1988). But once the hydraulic threshold of laminar to turbulent flow has been passed, cave enlargement proceeds more rapidly. Cave passages to 3 m diameter can develop in about 10^4–10^5 years.

The development of subterranean karst hydrology also depends on the manner in which water enters the karst. Rainwater that falls directly onto the limestone outcrop is known as *autogenic* recharge. It infiltrates diffusely into the rock via countless fissures. By contrast, rain that falls onto impervious non-karstic rocks but later flows onto the karst is known as *allogenic* recharge. It runs off as an organized stream, which sinks underground at the end of a BLIND VALLEY soon after encountering the limestone. Places where streams disappear underground are called *swallow holes*, *swallets*, *stream-sinks* or *ponors*. Sites where water cascades steeply into open pits are sometimes called pot-holes or abîmes (French).

Autogenic waters are mainly acidified by carbonic acid, but allogenic waters may have flowed from peat bogs (and hence contain organic acids) or have encountered sulphide minerals when draining from shales and hence contain sulphurous acid. They also tend to have a greater mechanical load, which can abrade limestone and help to incise cave floors. Streams sinking along the allogenic input boundary converge underground and emerge at springs at the output boundary of the system, thereby establishing dendritic subterranean drainage networks. The flow in these conduits is turbulent, as opposed to the much slower flow in tiny interconnecting pores and fissures which is laminar. Processes of CAVE network development (*speleogenesis*) are discussed by White (1988), Ford and Williams (1989), Gillieson (1996) and Klimchouk *et al.* (2000).

Some karsts have very extensive cave systems. The longest in the world is the Mammoth–Flint Ridge–Roppel–Procter System in Kentucky, USA. It comprises an interconnected system of essentially horizontal dendritic passages at different levels, totalling over 530 km in length, but developed in only about 100 m vertical stratigraphic thickness of limestone. The world's deepest known caves are Voronja Cave, Arabika massif, West Caucasus, at over 1,750 m deep, Reseau Jean Bernard (1,602 m) in France and Lamprechtsofen–Vogelschacht (1,535 m) in Austria.

Caves can be some of the world's oldest landforms, because they are located deep underground and so are protected for a long time from surface

denudation. Thus sediments in the Mammoth Cave system have been dated by radioactive decay of cosmogenic isotopes to 3.5 Ma (Granger *et al.* 2001), and fossil hominid and animal remains in cave sediments in South Africa have been dated by palaeomagnetism to a similar age.

Once the input–output connections are established diffuse autogenic recharge infiltrates the bedrock. Most of the dissolutional attack takes place on bedrock just beneath the soil, close to where CO_2 is generated and consequently where the percolating water attains its greatest aggressivity. Thus up to about 90 per cent of the corrosion is accomplished in the top 10 m or so of the limestone outcrop. Since water penetrates underground mainly by means of joints and faults, these fissures become more widened by corrosion near the surface than they are at greater depth. The surface of karst is therefore very permeable, but permeability (the capacity to transmit water) decreases with depth. This highly corroded superfical zone is termed the EPIKARST (or alternatively the *subcutaneous zone*).

Before rainwater drains underground it flows across rocky outcrops on the surface. The resulting corrosion of these outcrops yields a small-scale solution sculpture of vertically fluted rock and widely opened joints, collectively known as *lapiés* (French) or *Karren* (German). Rocky spires produced in this way can sometimes be tens of metres high, although individual KARREN are normally smaller. These forms are particularly common above the tree-line where soil and vegetation are thin or absent, but karren also develop beneath a soil and vegetation cover. When the bedrock has been scoured by glaciation and stripped of loose debris, postglacial weathering produces bedding-plane surfaces with joints opened by dissolution. In northern England the whole surface is known as a LIMESTONE PAVEMENT, with the widened joints called *grikes* and the intervening blocks called *clints*. Large solutionally widened joint corridors are known as *bogaz* and complex networks of such features are sometimes known as labyrinth karst. Areally extensive expanses of any or all of these features, especially above the tree-line, constitute a *karrenfeld*.

Sometimes part of the carbonate dissolved near the surface is precipitated further down in pores and fissures in the bedrock. This is particularly common in porous limestones such as coral, and results in several metres of the bedrock near the surface being hardened by being made less permeable, a process known as CASE HARDENING. This is encouraged in hot climates in particular by evaporation near the surface. Carbonate crusts produced by secondary precipitation of carbonate are known as CALCRETE or *caliche*. Sometimes induration of the rock by case hardening proceeds in calcareous dune limestones (AEOLIANITE) at the same time as karst development is occurring. This produces a style of landscape known as SYNGENETIC KARST. After further percolation the seepage waters may emerge into cave passages. Since cave air usually has CO_2 concentrations similar to the open atmosphere, the emergence of supersaturated percolation waters results in CO_2 degassing and in the precipitation of calcite in the form of stalactites, stalagmites and flowstones (collectively termed SPELEOTHEMS).

Whereas the most characteristic subterranean features of karst are caves, the most typical surface landforms are closed depressions, especially DOLINES, which are enclosed bowl or saucer-shaped hollows, usually of a few hundred metres in diameter and some tens of metres deep. When dolines occupy all the available space, the surface has a relief like an egg-tray and is known as *polygonal karst*, but this does not always develop and often dolines are dispersed or in clusters across an undulating surface. Polygonal karsts can have doline densities ranging from 4–55 per square kilometre. The particularly large and correspondingly deep solution dolines of some tropical and subtropical karsts are also called *cockpits*, a Jamaican term.

Solution dolines (Plate 68) develop in the subcutaneous zone and drain water centripetally to enlarged fissures that discharge it vertically to the deep groundwater system (Figure 97). Small

Plate 68 The hallmark landform of karst: solution dolines near Waitomo, New Zealand

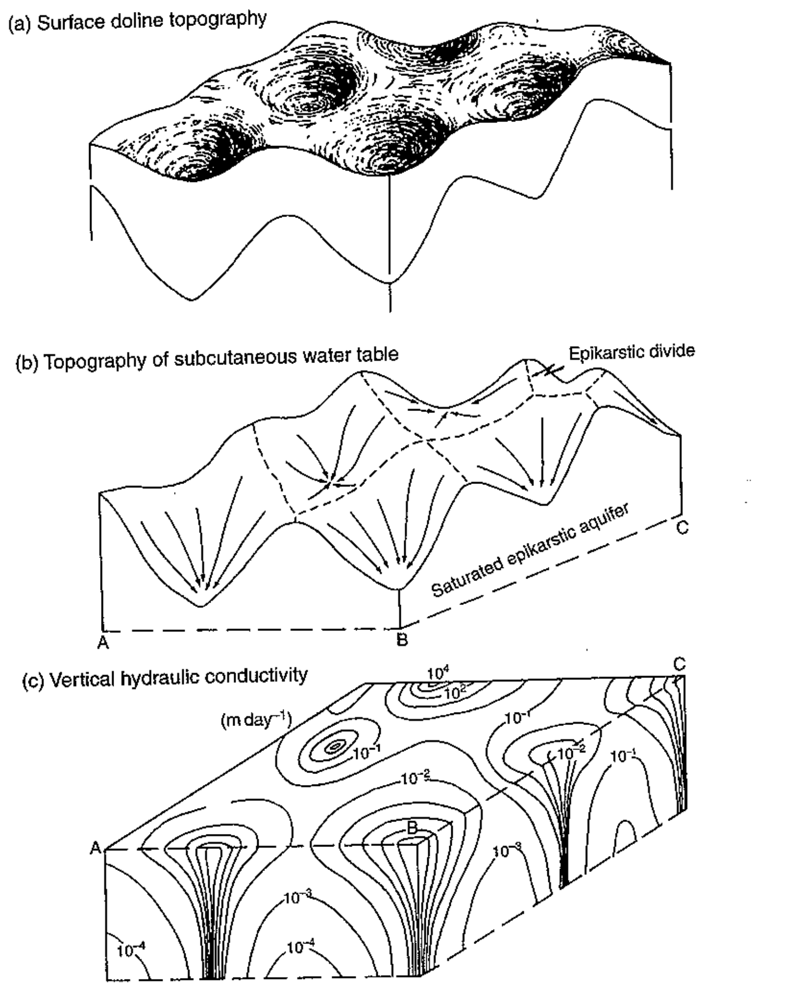

Figure 97 A model showing the relationship of solution doline development to flow paths in the epikarst (subcutaneous zone) and vertical hydraulic conductivity (from Williams 1985.)

allogenic streams also form enclosed basins, where they disappear underground. Large allogenic streams penetrate further into the karst in well-defined valleys before they sink, and produce landforms known as BLIND VALLEYS, because their valleys usually terminate abruptly in a cliff or steep slope. The sinking streams give rise to caves, and if a cave roof is close to the surface it sometimes collapses, producing a cylindrical or crater-like depression termed a *collapse doline*. A collapse that exposes an underground river is sometimes called a *karst window*. Some caves can be completely unroofed by progressive collapse, producing a gorge of cavern collapse (though not all gorges in karst are produced in this way, many being produced by antecedent drainage). Where

doline collapse intersects the water table, the steep-sided enclosed depression holds a lake, such features being called CENOTES in the Yucatan Peninsula.

As dolines evolve, they often enlarge laterally and coalesce, producing compound closed depressions known as *uvalas*. Where the rate of vertical incision of dolines is significantly greater than the rate of solutional denudation of the intervening land, the inter-doline areas develop into hills. This is particularly common in humid tropical and subtropical karsts, where residual hills can be so well developed that they visually dominate the landscape, giving rise to a style of landscape called *cone karst*. In China, *fengcong* is the term used to describe such karsts (Plate 69).

Many karst areas have developed on rocks that have been folded and faulted. These tectonic influences considerably complicate karst evolution and are of major significance in guiding

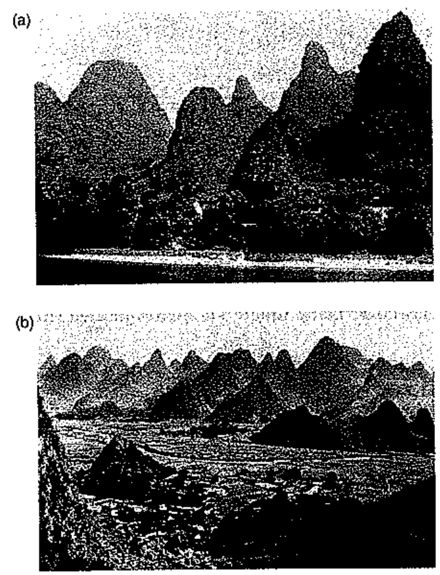

Plate 69 Two examples of humid subtropical karst in Guangxi province, China. The skylines are dominated by the conical forms of the hills, but enclosed depressions occur between them. When depression floors reach the water table they widen at their base, isolate the hills and extend the floodplain surface

groundwater flow and denudation of the surface. Faulted terrains often provide the conditions in which the largest enclosed karst depressions – known as POLJES – are developed, some exceeding 100 km^2 in area. The term 'polje' is of Slav origin and means 'field', probably because it was the largest area of flat tillable land in the karst. Many examples of these features are found in the Dinaric karst, where poljes are often located in faulted basins. Ford and Williams (1989) define three types: border polje, structural polje and baselevel polje – according to the dominant influences in their evolution (Figure 98).

Sometimes a particularly large blind valley encloses a basin of a square kilometre or more with a well-developed flat, floodplain floor, and it may receive more than one allogenic stream sinking at different points. Such large enclosed depressions at the edge of karsts are known as *border poljes*. Because of their relationship to sinking streams the floors of poljes often flood, particularly when the discharge of the inflowing river is greater than can be absorbed by the stream-sink(s). There is no clear demarcation between blind valleys and border poljes. They are transitional forms, the larger ones with particularly flood-prone flattish floors being called poljes.

Poljes may also be found in the interior of karsts, where structural dislocations have produced tectonic depressions with inliers of relatively impervious rocks. In these cases, the inlier acts as a dam on regional groundwater movement, forcing it to emerge as springs on the upstream side of the barrier. Water then flows across the impermeable inlier, to sink in ponors on the downstream side, the intervening region being developed into an alluviated plain. These features are known as *structural poljes*.

Genetically distinct from the above is the *base-level polje*, which is a very large enclosed depression entirely in karst rock that has been incised by solution down to the level of the *epiphreatic* zone (the zone of fluctuation of the water table). Such poljes are typically located close to the outflow boundary of a karst. They have swampy floors and can be envisaged as windows on the water table. Hence they inundate when the regional water table (or piezometric surface) rises in the wet season.

When vertical denudation eventually reduces the bottom of closed depressions to the level of the regional water table, they can incise no further, so instead they widen their floors. As a result residual hills between dolines become isolated, such

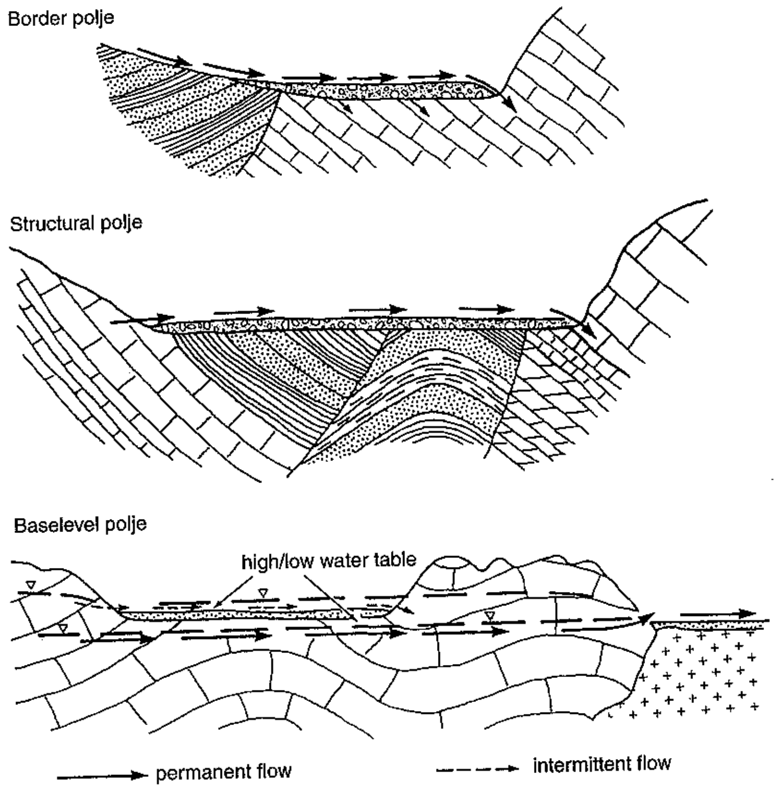

Figure 98 Three main types of polje. These are the largest enclosed depressions found in karst landscapes. Their flat floors are prone to flood and may cover many square kilometres in area (from Ford and Williams 1989.)

landforms being called *hums* in Europe. The lower slopes of such hills, which are usually of a rounded conical shape, can be over-steepened by undercutting and may collapse at their base, a process brought about by the corrosional attack of swamp waters – made particularly vigorous if allogenic rivers periodically flood the intervening plains. This is especially common in tropical humid karsts, where landscapes of steep isolated hills are referred to as *tower karst* (*Kegelkarst*, German), superb examples being known in southern China (where it is called *fenglin*). In the Caribbean isolated karst hills produced in this way are called *mogotes*.

During the end stages of karst denudation caves are drained and dismembered and their remnant passages are left at various elevations within residual karst hills, until eventually even the residual hills are removed by solution and only a corrosion plain is left. A superb example of a corrosion plain is the Gort lowland of counties Clare and Galway in western Ireland, where Pleistocene glaciations have stripped away the mantle of residual soil, alluvium and loose rock to reveal the karstified bedrock beneath.

Uplift can rejuvenate karst systems. But whereas in the first cycle of karstification the rock was unweathered and had only primary porosity, in the

second cycle there is an inheritance of landforms on the surface and secondary porosity underground. Thus a new phase of karst evolution would exploit the inherited features and develop them further.

Modelling karst evolution

Various attempts have been made to model karst landscape development. Early conceptual models of karst evolution were presented by Grund (1914) and Cvijić (1918) (see translations into English in Sweeting 1981), but it was not until the late twentieth century that models became quantitative.

Ford and Ewers (1978) used a physical laboratory model to elucidate the development of protocaves and successive flow paths. White (1984) developed a theoretical expression showing the relationship between chemical and environmental factors in the solutional denudation of limestones and he also developed a model of the development of cave passages (White 1988).

Ahnert and Williams (1997) developed a 3-dimensional model of surface karst landform development that started with a terrain in which proto-conduit connections were already established and then showed how the relief might develop given different assumptions about starting conditions, such as randomly disposed sites of greater permeability or random variations in initial relief (Figure 99). This model illustrates sequential steps in the development of doline and polygonal karst and reveals the importance of divergent and convergent flow paths in explaining the development of residual cones between incising depressions.

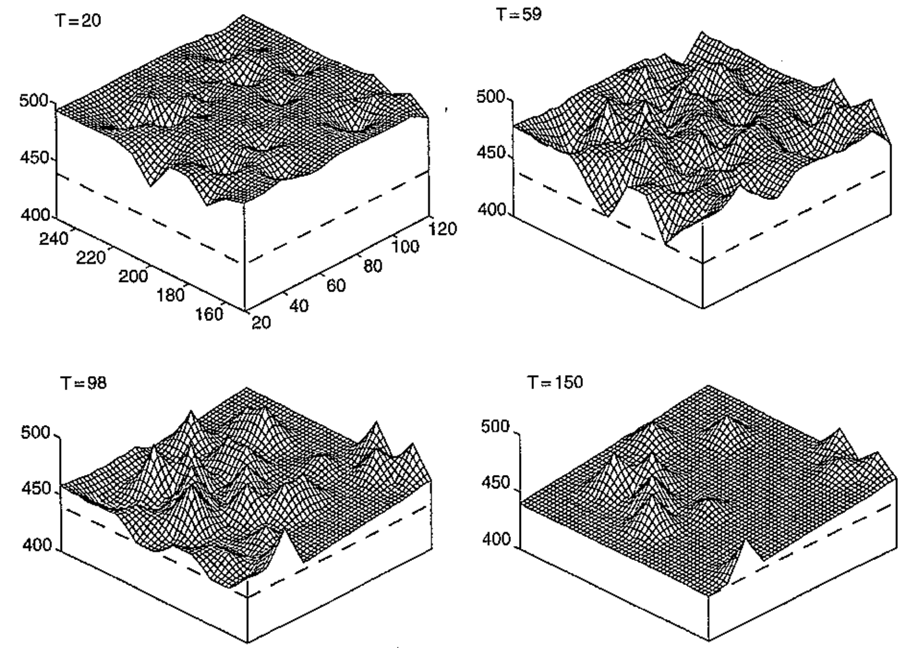

Figure 99 An illustration from one run of a three-dimensional theoretical model of karst landform development. The model shows an undulating surface with dolines at time 20 (T = 20), the development of polygonal karst by T = 59 (when some doline bottoms attain base level (shown by dashed line), the commencement of isolation of residual hills by T = 98, and the development of a corrosion plain with isolated hills by T = 150. The corrosion plain has a gradient towards the left because of the slope (hydraulic gradient) of the water-table (from Ahnert and Williams, 1997.)

Other more recent models associated with dissolution and the processes governing the evolution of karst are presented in Klimchouk *et al.* (2000) and Gabrovšek (2002).

References

Ahnert, F. and Williams, P.W. (1997) Karst landform development in a three-dimensional theoretical model, *Zeitschrift für Geomorphologie, Supplementband* 108, 63–80.

Cvijić, J. (1960) *La Géographie des Terrains Calcaire*, Monographies tome 341, No. 26, Belgrade: Academie Serbe des Sciences et des Arts.

Ford, D.C. and Ewers, R.O. (1978) The development of limestone cave systems in the dimensions of length and depth, *Canadian Journal of Earth Science* 15, 1,783–1,798.

Ford, D.C. and Williams, P.W. (1989) *Karst Geomorphology and Hydrology*, London: Chapman and Hall.

Gabrovšek, F. (ed.) (2002) *Evolution of Karst: from Prekarst to Cessation*, Postojna–Ljubljana, Inštitut za raziskovanje krasa, ZRC SAZU, Založba ZRC (Zbirka Carsologica).

Gillieson, D. (1996) *Caves: Processes, Development, Management*, Oxford: Blackwell.

Granger, D.E., Fabel, D. and Palmer, A.N. (2001) Pliocene-Pleistocene incision of the Green River, Kentucky, determined from radioactive decay of cosmogenic ^{26}Al and ^{10}Be in Mammoth Cave sediments, *Geological Society of America Bulletin* 113(7), 825–836.

Klimchouk, A.B., Ford, D.C., Palmer, A.N. and Dreybrodt, W. (eds) (2000) *Speleogenesis: Evolution of Karst Aquifers*, Huntsville, AL: National Speleological Society Inc.

Sweeting, M.M. (ed.) (1981) *Karst Geomorphology*, Benchmark Papers in Geology 59, Stroudsburg, PA: Hutchinson Ross.

White, W.B. (1984) Rate processes: chemical kinetics and karst landform development, in R.G. LaFleur (ed.) *Groundwater as a Geomorphic Agent*, 227–248, Boston: Allen and Unwin.

——(1988) *Geomorphology and Hydrology of Carbonate Terrains*, Oxford: Oxford University Press.

Williams, P.W. (1985) Subcutaneous hydrology and the development of doline and cockpit karst, *Zeitschrift für Geomorphologie*, 29, 463–482.

——(2003) Karst evolution, in J. Gunn (ed.) *Encyclopedia of Caves and Karst*, New York: Routledge.

PAUL W. WILLIAMS

KETTLE AND KETTLE HOLE

Kettle holes are depressions formed by the melt of discrete blocks of glacier ice that have been partially or completely buried by GLACIFLUVIAL sediments. Kettle holes have been reported from many present-day proglacial environments. Ice blocks originate in three ways: (1) detachment from the glacier snout due to ablation (Rich 1943), (2) transport on to the outwash plain or 'sandur' by glacial streams or rivers (Maizels 1977) and (3) release and transport on to the sandur surface during jökulhlaups (glacier OUTBURST FLOODS) (Fay 2002). Kettled or pitted sandur is glacial outwash in which numerous kettle holes have formed.

Kettle holes may be inverse-conical or steep-walled in shape. Inverse-conical kettles form due to the melt of partially or totally buried ice blocks and develop by sediment slumping and avalanching down the kettle walls. Inverse-conical kettles formed by the melt of partially buried blocks may possess raised diamict (see DIAMICTITE) rims (rimmed kettles) and/or diamict mounds in the base of the depression. Steep-walled kettles form by collapse of overlying sediment into voids created by the *in situ* melt of completely buried ice blocks. Ice blocks transported on to the sandur by water are often progressively buried by glaciofluvial sediments. However, during a jökulhlaup, sediment deposition can be so rapid that small ice blocks are incorporated within the flow and deposited simultaneously with flood sediments.

The physical properties of the sediment in which kettle holes form control their collapse sequence (Fay 2002). Shallow kettles with vertical or inward-dipping walls, whose base is a coherent block of sediment, form in coarse, clast-supported sediments. In coarse sediments dominated by matrix-support or in entirely fine-grained sediments, deeper kettles with steeply dipping or overhanging walls form, often through sudden roof collapse. Steep-walled kettle holes may form over small, buried blocks, or over larger, buried ice blocks which melt irregularly. Steep-walled kettles can develop into inverse-conical kettle holes by slide or avalanche of sediment into the kettle hollow.

Since the development of kettle holes is similar over both stagnant glacier ice and flood-related ice blocks, it may be difficult to differentiate between flood-related kettles and kettles produced by non-fluvially driven processes. However, on a palaeoflood surface, a flood origin for kettle holes is indicated by a distinct radial pattern of kettle holes on prominent outwash fans reflecting flow expansion, and/or a decrease in kettle size down sandur relating to a progressive decrease in stream power.

References

Fay, H. (2002) Formation of kettle holes following a glacial outburst flood (jökulhlaup), Skeiðarársandur, southern Iceland, in A. Snorrason, H.P. Finnsdóttir and M. Moss (eds) *The Extremes of the Extremes: Extraordinary Floods*, IAHS Publication 271, 205–210.

Maizels, J.K. (1977) Experiments on the origin of kettle-holes, *Journal of Glaciology* 18, 291–303.

Rich, J.L. (1943) Buried stagnant ice as a normal product of a progressively retreating glacier in a hilly region, *American Journal of Science* 241, 95–99.

Further reading

Maizels, J.K. (1992) Boulder ring structures produced during jökulhlaup flows - origin and hydraulic significance, *Geografiska Annaler* 74A, 21–33.

HELEN FAY

KNICKPOINT

Knickpoint (nickpoint) refers to a substantially steepened section of a stream long profile. In cyclic interpretations of landscape (Davis 1899) knickpoints carried a new 'cycle' of erosion inland replacing an older, more elevated low relief surface and producing a gorge (see GORGE AND RAVINE). For knickpoint erosion in alluvial material, soil and rill systems, and in the laboratory, the term 'headcut' is used (Bennett 1999), or for extended steep reaches, knickzone (Downs and Simon 2001). The term WATERFALL is synonymous with knickpoint. The steepened valley walls below the knickpoint subsequently collapse by mechanical failure following stress release (Philbrick 1970), enhanced groundwater drainage, and by stream erosion undercutting the valley wall. Despite morphological similarities between knickpoints in cohesive clay materials and those in bedrock, they may not retreat by similar processes. Both cases depend upon the headwall being continually re-steepened with the bulk of any detrital material present at the base of the knickpoint being removed. In weak material incapable of maintaining a cap to the headwall the knickpoint face may rotate as it retreats: longitudinal slope diminishing until the knickpoint is removed.

Niagara Falls has a mean flow rate (before abstraction for power generation) of $5{,}730\,m^3\,s^{-1}$. It retreated at a rate $1.5\,m\,yr^{-1}$ in the last decades of the nineteenth century (Philbrick 1970); a figure consistent with the estimated rate for the entire postglacial period (14,000 calendar years). When flow was only 10 per cent of the present value (Tinkler *et al.* 1994) for 5,000 years, the recession rate was between 0.10 and $0.15\,m\,yr^{-1}$, and with a reduced elevation for the headwall. Although $1.5\,m\,yr^{-1}$ is high, the recession rate of the 12 m St Anthony Falls on the Missouri may have been comparable in the nineteenth century (Winchell 1878; Sardeson 1908). The inference for both waterfalls is that the headwall retreated in an essentially parallel fashion during periods of thousands of years.

Recession rates for other major knickpoints are known only for a few systems at present. In south-east eastern Australia (van der Beek *et al.* 2001) rates of the order of 1,000 m per million years have been estimated, and half the rate that has been reported in adjacent areas of Australia (Nott *et al.* 1996; Seidl *et al.* 1997). In southern Africa Derricourt (1976) has suggested rates for the recession of the Batoka Gorge, below the Victoria Falls, have varied between 0.09 and 0.15 $m\,yr^{-1}$ over the last million years or so. Stranded and stationary surface knickpoints have been reported in karst terrains (Fabel *et al.* 1996; Youping and Fusheng 1997).

The process of knickpoint recession is far from clear. There is no published evidence that for large waterfalls headwall undercutting takes place in the plunge pool. In non-cohesive sediments with low slope the entire headcut may be submerged but this has only rarely been described for large bedrock forms (Rashleigh 1935). More probably headwall recession above the plunge pool level proceeds by subaerial weathering in a very damp environment; by sapping, from water fed to the vertical face from the upper river bed (Krajewski and Liberty 1981); and by stress release close to rock face weakening joints and bedding plains prior to block release. In the upper caprock zone (if present) accelerating water approaching the waterfall edge eventually exerts enough force to tear blocks out of the undermined capping beds (Philbrick 1970). At full flow, on waterfall faces less than vertical, erosional wear by water and entrained sediment may effect rock removal (Bishop and Goldrick 1992; and examples in Rashleigh 1935). Knickpoints in alluvial sediments can progress much faster (Simon and Thomas 2002) cite rates of 0.7 to $12\,m\,yr^{-1}$.

References

Bennett, J.S. (1999) Effect of slope on the growth and migration of headcuts in rills, *Geomorphology* 30, 273–290.
Bishop, P. and Goldrick, G. (1992) Morphology, processes and evolution of two waterfalls near Cowra, New South Wales, *Australian Geographer* 23, 116–121.
Davis, W.M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.
Derricourt, R.M. (1976) Retrogression rate of the Victoria Falls and the Batoka Gorge, *Nature* 264, 23–25.
Downs, P.W. and Simon, A (2001) Fluvial geomorphological analysis of the recruitment of large woody debris in the Yalobusha River network, Central Mississippi, USA, *Geomorphology* 23(1–2), 65–91.
Fabel, D., Henricksen, D. and Finlayson, B.L. (1996) Nickpoint recession in karst terrains: an example from the Buchan karst, southeastern Australia, *Earth Surface Processes and Landforms* 21, 453–466.
Gregory, J.W. (1911) Constructive waterfalls, *Scottish Geographical Magazine* 27, 537–546.
Krajewski, J.L. and Liberty, B.A. (1981) Present dynamics of Niagara, in I. Tesmer and J.C. Bastedo (eds) *Colossal Cataract: Geological History of Niagara Falls*, 63–93, Albany: University of Toronto, State University of New York.
Nott, J., Young, R. and McDougall, I. (1996) Wearing down, wearing back, and gorge extension in the long-term denudation of a highland mass: quantitative evidence from the Shoalhaven catchment, southeast Australia, *Journal of Geology* 104, 224–232.
Philbrick, S.S. (1970) Horizontal configuration and the rate of erosion of Niagara Falls, *Geological Society of America Bulletin* 81, 3,723–3,732.
Rashleigh, E.C. (1935) *Among the Waterfalls of the World*, London: Jarrolds.
Sardeson, F.W. (1908) Beginning and recession of Saint Anthony Falls, *Geological Society of America Bulletin* 19, 29–52.
Seidl, M.A., Finkel, R.C., Caffee, M.W., Hudson, G.B. and Dietrich, W.E. (1997) Cosmogenic isotope analyses applied to river longitudinal profile evolution: problems and interpretations, *Earth Surface Processes and Landforms* 22, 195–209.
Simon, A. and Thomas, A. (2002) Processes and forms of an unstable alluvial system with resistant, cohesive streambeds, *Earth Surface Processes and Landforms* 27(7), 699–718.
Tinkler, K.J., Pengelly, J.W., Parkins, W.G. and Asselin, G. (1994) Postglacial recession of Niagara Falls in relation to the Great Lakes, *Quaternary Research* 42, 20–29.
van der Beek, P., Pulford, A. and Braun, J. (2001) Cenozoic landscape development in the Blue Mountains (SE Australia): lithological and tectonic controls on rifted margin morphology, *Journal of Geology* 109, 35–56.
Winchell, N.H. (1878) The recession of the Falls of St Anthony, *Journal of the Geological Society* 34, 886–901.
Youping, T. and Fusheng, H. (1997) Research on waterfall calcareous tufa mats from Xiangzhigou, Guizhou, *Carsologica Sinica* 16(2), 145–154.

Further reading

Wohl, E. (2000) *Mountain Rivers*, Washington, DC: American Geophysical Union.

KEITH J. TINKLER

L

LAGOON, COASTAL

There are three main meanings to the term lagoon. The most usual is to describe a stretch of salt water separated from the sea by a low sandbank or coral reef. Another meaning is that for a small freshwater lake near a larger lake or river. The third usage is for an artificial pool used for the treatment of effluent or to accommodate an overspill from surface drains. This entry is soley concerned with the first of these meanings.

Coastal lagoons are mostly to some degree estuarine, are usually shallow, and have generally been partly or wholly sealed off from the sea by the deposition of spits or barriers, by localized tectonic subsidence, or by the growth of coral reefs. They range in size from over 10,000 km^2 to less than 1 ha. They occur on about 12 per cent of the length of the world's coastline (Bird 2000). They are best formed on transgressive coasts, particularly where the continental margin has a low gradient and sea-level rise is slow. They are ephemeral features and their depths and areas become reduced by sedimentation from inflowing rivers, as well as by accumulation of sediment washed in from the sea, wind-blown material, and chemical and organic deposits. Indeed, they can be classified on the basis of whether they are infilling or increasing in size (Nichols 1989). The infill of some lagoons, particularly those that are parallel to the shore, may involve the development of cuspate divisions that divide the lagoon into a series of segments. These divisions have been attributed *inter alia* to winds blowing along the length of the lagoon producing waves which build spits that isolate the lagoon into separate basins.

The entrances between lagoons and the sea vary in origin and form. Some are residual gaps that persisted between spits or barrier islands where the lagoon was never completely sealed off from the sea. Others are caused by breaching either by storm waves or by floods from on land. Their configuration is the outcome of a contest between the inflow and outflow of currents, which keeps entrances open, and the effects of onshore and longshore drift of sediment, which tends to seal them off. Lagoon entrances tend to be larger, more numerous and more persistent on barrier coastlines where relatively large tidal ranges generate strong currents.

A good review of the diversity of coastal lagoon morphologies and evolution is provided by Cooper (1994).

References

Bird, E.C.F. (2000) *Coastal Geomorphology: An Introduction*, Chichester: Wiley.
Cooper, J.A.G. (1994) Lagoons and microtidal coasts, in R.W.G. Carter and C.D. Woodroffe (eds) *Coastal Evolution*, 219–265, Cambridge: Cambridge University Press.
Nichols, M.M. (1989) Sediment accumulation rates and sea-level rise in lagoons, *Marine Geology* 88, 201–219.

A.S. GOUDIE

LAHAR

An Indonesian word originally used by Escher (1922) for a hot volcanic mudflow that had been generated by an eruption through a crater lake. The term specifically applied to the 1919 volcanic mudflows originating from the crater lake of Mt Kelut, on Java, which inundated over 130 km^2 of the surrounding lowlands, with the loss of over 5,000 lives. The word quickly gained acceptance as a general term for a mudflow on the flanks of a volcano, irrespective of its triggering mechanism.

With increased understanding of water-saturated flow processes, the term has evolved to include both mudflows and debris flows, and later hyperconcentrated flows, on the flanks of a volcano. This has led to the more generally accepted definition: a rapidly flowing mixture of rock debris and water (other than normal stream flow) from a volcano (Smith and Fritz 1989). This definition describes the flow but not the resultant deposit.

Lahars are denser than normal stream flow because of their high sediment loads, causing them to move faster due to greater gravity forces and damped internal energy losses. The highly concentrated slurry usually exhibits a yield strength, behaving like wet cement but also exhibiting features of cohesionless grain flow. At lowering velocities, high concentration lahars may quickly halt, 'freezing' the coarser fraction within its finer matrix as a single massive sedimentary unit. The matrix strength, buoyancy and grain-dispersive pressures help support the coarser particles (Iverson 1997). Thus, resultant deposits are poorly sorted and may show reverse or no grading because little or no time was available for settling to occur. Thin, fine-grained sole-layers are attributed to shear and cataclasis within the basal flow. Lahars may travel as successive flow surges, which in high sediment concentration flows tend to continually shunt and overtop previous deposits. Clay-poor lahars may exhibit a frontal cliff or snout and show coarse bouldery levees at their margins.

A feature of lahars is 'bulking', when on the steep slopes of a volcano the flow typically erodes loose sediment over which it is flowing to increase its volume several times. On lower gradient slopes the reverse may happen leading to progressive dilution of the flow and formation of a HYPERCONCENTRATED FLOW or a normal flood (Pierson 1998). A water wave ahead of the lahar has been observed in some instances, which may represent a solitary wave or soliton (Cronin *et al.* 2000) or may be caused by under-ramping of the dense flow behind (Manville *et al.* 2000).

Lahars, armoured with their coarse bouldery loads, are a devastating volcanic hazard, capable of killing large numbers of people and removing all structures in their path. In recorded history at least 64,000 persons have been killed by lahars (Neall 1996). Lahars may vary in size from small volume events (<0.1 million m^3) to large-scale collapses of volcanic edifices ($>3{,}500$ million m^3). Measured peak discharges of historical lahars range from 100 to $>100{,}000\,m^3\,s^{-1}$. Lahars are capable of travelling up to 150 km/h and may travel for hundreds of kilometres down valleys. This mobility is attributed to positive pore-fluid pressures, which greatly decrease internal friction within the flow (Hampton 1979). Hence the most hazardous lahar zones are adjoining river courses draining volcanoes, particularly those draining crater lakes.

Lahars may be a more common hazard at stratovolcanoes than the geological record attests. Of twenty lahars observed in the Whangaehu catchment at Mt Ruapehu, New Zealand, during 1995, only one resultant deposit is preserved. This highlights the preservation potential of only larger events.

Lahars may be generated by both eruptive and non-eruptive mechanisms. Lahar triggering eruptive mechanisms include phreatic or phreatomagmatic explosions (often from hydrothermally altered and structurally weakened edifices), displacement of waters by eruptions through a crater lake, pyroclastic flows admixing with water in rivers or lakes, and volcanic melting of snow and ice. Non-eruptive triggering mechanisms include collapse of the walls of crater lakes, and heavy rains falling on recently erupted materials.

In the most recent large-scale volcanic disaster of modern times more than 23,000 persons were killed by lahars from Nevado del Ruiz, Colombia, in 1985. Relatively small eruptions at the summit produced pyroclastic surges, which quickly gouged and melted $0.06\,km^3$ of snow and ice to form lahars peaking at $48{,}000\,m^3\,s^{-1}$ and totalling 40–60 million m^3.

Large and extensive prehistoric lahar deposits are reported from many stratovolcanoes. One of the first to be recognized was the Osceola Mudflow from Mt Rainier, Washington State, USA, which filled proximal valleys 85–200 m deep before spreading over $350\,km^2$ of the Puget Sound lowlands about 5,600 years ago.

Mitigative measures to reduce damage from lahars include adequate real-time warning systems (such as acoustic flow monitors), reducing the level of water in crater lakes by engineering methods (such as tunnels at Mt Kelut, Indonesia), dams to reduce gradient and encourage sediment deposition, reduction of lake levels in hydro dams to accommodate sudden inflows, construction of embankments to divert flow away from assets

at most risk, or as at Mt Pinatubo, Philippines, continually elevating houses above each successive lahar deposit.

References

Cronin, S., Neall, V., Lecointre, J. and Palmer, A. (2000) Dynamic interactions between lahars and stream flow; a case study from Ruapehu volcano, New Zealand: reply, *Geological Society of America Bulletin* 112, 1,151–1,152.

Escher, B.G. (1922) On the hot 'lahar' (mud flow) of the Valley of Ten Thousand Smokes (Alaska), *Proceedings Koninklijke Akademie van Wetenschappen, Amsterdam* 24, 282–293.

Hampton, M.A. (1979) Buoyancy in debris flows, *Journal of Sedimentary Petrology* 45, 753–758.

Iverson, R.M. (1997) The physics of debris flows, *Reviews of Geophysics* 35, 245–296.

Manville, V., White, J.D.L. and Hodgson, K.A. (2000) Dynamic interactions between lahars and stream flow; a case study from Ruapehu volcano, New Zealand; discussion, *Geological Society of America Bulletin* 112, 1,149–1,151.

Neall, V.E. (1996) Hydrological disasters associated with volcanoes, in V.P. Singh (ed.) *Hydrology of Disasters*, 395–425, Dordrecht: Kluwer.

Pierson, T.C. (1998) An empirical method for estimating travel times for wet volcanic mass flows, *Bulletin of Volcanology* 60, 98–109.

Smith, G.A. and Fritz, W.J. (1989) Volcanic influences on terrestrial sedimentation, *Geology* 17, 375–376.

VINCENT E. NEALL

LAKE

Lakes are defined as bodies of slow moving water surrounded by land. They represent approximately 2 per cent of the Earth's surface but contain only about 0.01 per cent of the world's water (Wetzel 2001). While they are temporary features of the landscape on a geological timescale, they can exist for very long periods and therefore strongly influence the human development of a region. They can also provide a record of the region's environmental history in their sedimentary record. The study of lakes is called limnology and limnologists characterize lakes in a number of ways, including their geologic origin, mixing behaviour and nutrient status. While these classifications appear to be distinctly disciplinary, the geology, physics and chemistry of lakes interrelate significantly and all act to regulate the biological dynamics in lakes.

The depressions, or basins in the Earth's surface, that collect water to become lakes can be formed by several geologic and geomorphic processes. Catastrophic geologic origins include tectonic activity and volcanism. The deepest of the Earth's lakes are caused by tectonic faulting, whereas the clearest of lakes are found in the craters of old volcanoes. The majority of the Earth's lakes (72 per cent) have been created by glaciation (Kalff 2002). High alpine cirque (see CIRQUE, GLACIAL) lakes, lowland KETTLES AND KETTLE HOLES and glacial ice-scour lakes are abundant in the regions once covered by ice. Other lakes are generated by the modification of drainage systems, or the impoundment of flowing waters by natural disasters such as landslides or human activity such as damming. Riverine or fluvial lakes, such as those developed on floodplains, deltas and blocked valleys, comprise 10 per cent of the world's lakes and are the dominant lake type in low latitudes (Kalff 2002). Chemical dissolution of rocks also generates basins that collect water. Wind and ocean shoreline processes create barriers which act to block fresh water while animals and meteorites cause terrestrial depressions creating specific lake types. Some lake basins are excavated by wind (see PAN). The fact that lakes are most often created by processes that have dominated that landscape results in lakes of similar origin being regionally grouped. This is why there seems to be a preponderance of Lake Districts around the world, where bodies of water that appear and behave similarly due to their shared origin are identified as a lake grouping (e.g. Cumbria Lakes, England; Finger Lakes, New York; Great Lakes, North America). Such groupings provide opportunities for regional-scale research and have facilitated the study of lake types and processes.

The shape of the lake basin, or its morphometry, is a function of the lake origin but over time it does change as sedimentation and shoreline infilling occurs as part of the natural succession process. Zonation in lakes is also a function of morphometry. The nearshore region of lakes is called the littoral zone while the open water deeper portion is called the pelagic zone. Littoral zones support rooted plants, some of which are visible above the water surface and are called emergent plants. The littoral extends to the depth of water able to support rooted plants and it represents that region to which light can penetrate, allowing photosynthesis of these primary producers. The littoral zone may be extensive in a shallow dish-shaped lake or it may represent only a small nearshore area if the slope is very steep (see Figure 100). Primary productivity (plant growth) in the littoral is usually greater than in the pelagic zone as it is more closely linked to the catchment from which it receives nutrients for plant growth.

Lakes are classified physically by their mixing behaviour which is a function of their thermal structure. Solar energy enters the lake at the water surface and is attenuated as it moves down through the water column. Heat is therefore transferred predominantly to the surface waters. Warm water lies above the colder, denser water creating a separation or stratification of water layers. The warmer surface waters are known as the epilimnion while the cooler bottom waters are called the hypolimnion. The point where the temperature transition is most extreme between these two layers is called the thermocline and that portion of the lake is termed the metalimnion (see Figure 100).

Water is an unusual liquid in that it is most dense at 4 °C, meaning that water both warmer and cooler than 4 °C will be more buoyant and rise to a layer above. Water in its solid form, ice, is less dense than water and thus floats at the lake surface. The unique density characteristics of water ensure that lakes do not freeze from the bottom up and that most temperate lakes have bottom waters with moderate under-ice temperatures (4 °C) to support life.

Depending on the temperature (and therefore density) differences between the epilimnion and hypolimnion the two parcels of water may be very resistant to mixing. During periods of strong stratification the two compartments of water do not interact, or mix, and therefore their exchange of materials is restricted. The movement of both soluble and low density particulate material entering a lake can be confined to surface waters if the density differences at the metalimnion are extreme. When denser particulate matter settles through the metalimnion and enters the hypolimnion it will be stored and/or decomposed, but until the stratification is reduced all these materials (e.g. nutrients, pollutants, particles) will remain in the bottom waters. Chemical changes associated with the decomposition of the particles, such as reduced oxygen levels, are now confined to the bottom waters as there is little exchange of water and chemicals across the thermocline. In well-stratified, productive lakes oxygen depletion or anoxia often occurs in the bottom waters due to the oxygen demand of the organic-rich sediments (see Figure 100). Anoxic conditions at the sediment–water interface also generate chemical alterations in the sediment, resulting in increased exchanges between sediments and overlying water.

When stratification breaks down, due for example to the cooling of surface waters with changing seasons, and the water column becomes isothermal (one temperature) there is little resistance to mixing. Wind energy at the water surface can mix the full water column which is called 'turnover'.

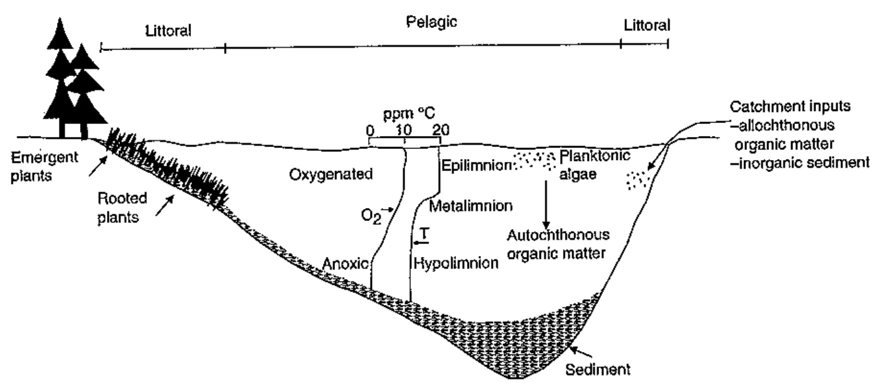

Figure 100 Limnological terms for the zonation, thermal structure and particulate sources to lakes. The general distribution of lake primary production and resultant oxygen profiles are also presented

This full lake mixing brings nutrients and chemicals from the hypolimnion to mix with the surface waters. Lakes in different climatic regions exhibit different annual patterns of stratification and so can be classified in this way. Lakes which mix twice a year are common in the temperate regions where cold winters give way to springtime heating generating isothermal conditions as the lake warms to 4 °C. Again in the autumn the lakes cool, causing destratification and an isothermal situation which allows autumnal winds to mix the bottom with the surface waters. This mixing, twice annually, is termed dimictic. Monomictic lakes mix only once per year and are found in high elevation and high latitude areas, while polymictic lakes, commonly located in equatorial regions, mix many times a year. Note that mixing is important in the transfer of both nutrients and pollutants in lakes. A less common but environmentally significant situation is a meromictic lake which has a bottom water layer that is chemically different from the rest of the water body. This generates a large enough density difference that even when the system is isothermal a chemical density gradient restricts this bottommost layer from mixing. This is often caused by salinity differences or groundwater inputs of differing chemistry. In this case the exchanges between the bottom sediments underlying the meromictic layer are restricted from delivering to the full water column.

Nutrient status is one variable used to chemically classify lakes. The terms oligotrophic, mesotrophic, eutrophic and hypereutrophic represent the spectrum of conditions from nutrient-poor through to nutrient-rich systems. In general, primary productivity increases across the spectrum with hypereutrophic lakes representing the stage where organic nutrient inputs (carbon, nitrogen, phosphorus) are high. Organic inputs to lakes can be derived from within the lake, termed autochthonous, or they can be delivered from the catchment or atmosphere, when they are called allochthonous. In-lake primary production of free-floating photosynthetic plants, or planktonic algae, and littoral rooted plants are the main sources of autochthonous organic matter. Catchment inputs from river inflows represent the main allochthonous contribution of organic matter. If the organic matter in lakes is not either eaten by an organism or decomposed during settling it will accumulate in the bottom sediments. Organic sedimentation rates increase over the trophic range as productivity increases.

The bottom sediments are also comprised of inorganics which have been eroded from the landscape and delivered via river inflows or shoreline erosion. As well, some autochthonous inorganics can be generated by chemical precipitation and biogenic processes. Sedimentation rates vary as a function of the location, size and activities in the catchment as well as the in-lake productivity, but in most natural temperate lakes ranges between 0.1 and 2 mm per year (Kalff 2002).

Globally much research has focused on the process and remediation of cultural eutrophication (Cooke *et al.* 1993). This accelerated change in tropic status occurs when unnaturally high phosphorus loads are received by the lake over relatively short time periods (decades). It is termed cultural eutrophication because the sources of the growth nutrient, phosphorus, are associated with human activities in the catchment, such as agriculture and sewage treatment. Increased primary productivity in both the littoral and pelagic zone, increased sedimentation rates, hypolimnetic oxygen depletion and the coarsening of fish species are often associated with anthropogenically induced alterations of tropic status. Remediation and management approaches for these problems are presented in Cooke *et al.* (1993).

As lakes are bodies of water that receive and store material from the surrounding catchments and the atmosphere they are of interest to geomorphologists as the accumulated sediment can reflect regional changes over time. Catchment erosion rates associated with changing land use, sediment source tracing, climatic variations, historic pollutant loads, flood records and vegetation patterns can be detected by evaluating different characteristics of the accumulated lake sediments. Sediment, collected by coring down through the accumulated material, can be horizontally sliced to differentiate sediments from specific time periods. Paleolimnology, or the use of lake sediments for reconstructing past events, requires some means of dating the accumulated material (see DATING METHODS) and a variety of methods exist (e.g. ^{210}Pb, ^{14}C, ^{137}Cs, thermoluminescence) but the precision and accuracy of each is restricted to specific time intervals. Given this, and the fact that lake sediments are temporally and spatially variable, it is very important to design the collection of cores and the analytical methods to suit the questions being addressed. Dearing and Foster (1993) provide a useful discussion of the problems, errors and implications of using sediment cores in geomorphic research. An earlier text by Hakanson and Jansson (1983) introduces the topic of lake sedimentology and provides information on physical, chemical and biological aspects of sediment.

Since 1970 the focus of limnological research has moved away from viewing the lake as a closed system upon which to do ecological research and more effort has been placed on linking catchment processes to lake conditions (Kalff 2002). Lakes are intimately connected to their catchments and therefore the role of lakes in geomorphological research and the role of geomorphologists in the interdisciplinary study of lakes are significant.

References

Cooke, G.D., Welch, E.B., Peterson, S.A. and Newroth, P.R. (1993) *Restoration and Management of Lakes and Reservoirs*, Boca Raton, FL: Lewis.

Dearing, J.A. and Foster, I.D.L. (1993) Lake sediments and geomorphological processes: some thoughts, in J. McManus and R.W. Duck (eds) *Geomorphology and Sedimentology of Lakes and Reservoirs*, 5–14, Chichester: Wiley.

Hakanson, L. and Jansson, M. (1983) *Lake Sedimentology*, Berlin: Springer-Verlag.

Kalff, J. (2002) *Limnology: Inland Water Systems*, Upper Saddle River, NJ: Prentice Hall.

Wetzel, R.G. (2001) *Limnology: Lake and River Ecosystems*, 3rd edition, Fort Worth: Academic Press.

Further reading

Hutchinson, G.E. (1957) *A Treatise on Limnology, Vol. 1. Geography, Physics and Chemistry*, New York: Wiley.

Vallentyne, J.R. (1974) *The Algal Bowl: Lakes and Man*, Special Publication 22, Department of the Environment Fisheries and Marine Service, Ottawa.

ELLEN L. PETTICREW

LAND SYSTEM

Land systems are an integral part of the land evaluation process. The first land systems studies were devised and carried out by Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) as part of their *Land Research Series* (Christian and Stewart 1968; Stewart 1968). A hierarchical classification of land was used. The smallest areal unit is the *land facet*, which is a relatively homogeneous area of land in terms of slope and soil; such as the midslope unit of a CATENA. Repetitive sequences of land facets form *land units*, which are the fundamental mapping unit of the land systems approach. Land units comprise a distinctive pattern of land facets (slopes, soils and vegetation), are essentially catenary in nature and classically were illustrated by block diagrams with explanatory annotations. *Land systems*, then, are larger areas comprising groups of land units representing recurring patterns of soils, topography and vegetation.

Work commenced in Australia in the late 1940s and continued to the 1970s, later refined by government agencies in some States. Similar surveys were carried out by the British Ministry of Overseas Development, particularly in Africa (Land Resources Development Centre 1966 onwards), and by the French scientific organization ORSTOM and the Netherlands International Institute for Aerial Survey and Earth Science (ITC) in many countries (Nossin 1977).

The objective was to provide rapid, cost-effective and objective assessments of land resources, initially for agriculture and forestry but more recently as frameworks for regional or national planning, in remote areas for which there was little if any base information. There were no maps depicting contours, soils or vegetation and little geological or climatic information for these areas. The primary data source was small-scale (1:80,000), monochrome aerial photography, flown primarily for military purposes from the 1940s. Thus the method of land systems surveys was essentially an air photo interpretation exercise, supported by ground truthing. Patterns of tone and texture were delimited on the aerial photographs. Offroad ground traverses were then designed to visit examples of each pattern on the ground in order to record the vegetation, soils, topography and geology. Since the tonal and textural patterns primarily reflect vegetation, plus areas of bare ground, rock outcrops and water bodies, vegetation associations were particularly important in characterizing land units and land systems.

However, several members of the early land systems teams were British-trained geomorphologists (including Mabbutt, Ollier, Twidale and Young). Not surprisingly, therefore, geomorphology played a large part in the written interpretations of the land systems. That geomorphology was heavily influenced by the then prevailing Davisian cycle of erosion which, as pointed out by Chorley (1965: 35–36) in a different context, has little if any bearing on the prediction of land suitability for agricultural or other land uses. This is also largely true of genetic interpretations of soils.

The first land system study in Australia was of the Katherine–Darwin region (Christian and Stewart 1953). As a result of the survey a relatively small area was identified as being possibly suitable for agriculture, and a research station was established to conduct detailed field investigations and experiments. A similar outcome was obtained from the East Kimberley survey. Both research stations are still operational, and in the East Kimberley the eventual outcome was the development of the Ord irrigation scheme. While the entire area of Papua New Guinea was eventually covered by land systems surveys, the spatial coverage of over twenty reports in the Land Research Series in Australia was more scattered. Debatably, the developments outlined above were the only non-educational, positive outcomes of the series of Land Research reports.

Weaknesses of the technique included: its reconnaissance-type of approach (which ironically was also its strength); its static coverage (processes were not included); its geomorphic base with a strong, denudation chronology flavour; its exclusive focus on the biophysical environment; and the crudity of the data. Nevertheless, land systems may provide a useful basis for later and more dynamically oriented work, especially when there is little background information on the environment or on ecological conditions. In Victoria, for example, geomorphic elements have been refined as part of a land systems review to provide a suitable framework for describing the spatial attributes of land (Rowan 1990).

Only in the Hunter Valley of New South Wales was a CSIRO land systems survey conducted in an area already well-developed for agriculture. The Hunter Valley survey covered a smaller area than previous surveys and that, plus the larger amount of information which was available for the region compared with the more remote regions, made it possible to provide greater detail (Story *et al.* 1963). But that survey also exposed the limitations of focusing on the biophysical environment. In most areas, planning for future land uses does not take place in unoccupied country but in places which are already occupied and the land already subject to land-use practices. Existing uses, therefore, must be incorporated in any survey designed to assist planning for future land uses.

The South Coast study

Realization of this self-evident truth led to the eventual abandonment by CSIRO of its limited land systems surveys (although the basic method is still used widely in various contexts), and the development of two separate but related lines of further work. One was a focus on databases, which eventually merged with the currently thriving area of research into and applications of Geographical Information Systems in land use planning; and the other was the *South Coast study* (also in New South Wales) (Basinski 1978). This was a very 'geographical' survey, in the sense that its approach would be familiar to any regional geographer. Not surprisingly, geographers were prominent amongst the team members – as physical geographers had been in the earlier land systems survey teams.

The South Coast study, a large, multidisciplinary project, conducted the familiar, integrated, land systems survey of the biophysical environment of the 6,000 km^2 region. Importantly, however, present land use, population demographics and socio-economic aspects such as settlements, transport networks, and the economy and social structure of the region were also included. The study explicitly set out to provide a 'rational basis for planning decisions on a wide variety of land uses', as well as investigating methods of providing and analysing biophysical and socio-economic data for the region (Basinski 1978). This survey was a precursor of 'land suitability' surveys and regional planning and, more particularly, of CSIRO's Siroplan approach to land suitability evaluation for integrated, regional land use planning (Cocks *et al.* 1983).

Recent approaches to land evaluation

The term 'land systems' continues to be used and modified in some jurisdictions, particularly by agricultural agencies (at least in Australia). But it is the mapping unit of the Land Research Series, the *land unit*, which provides the basis of several approaches; notably FAO's land suitability evaluation (FAO 1976) and land use planning frameworks (FAO 1993), as well as regional planning, at least in Western Australia, in the form of the basic planning and land management unit (example: WAPC 1996).

Improved technology and data availability have made possible considerable improvements in land evaluation methods and the quality of the output, and Davidson (2002) has provided a useful review of recent developments in the assessment of land resources. The availability of high resolution, satellite-borne remote sensing imagery, the rapid development of geographic information systems (or science) (GIS), and geostatistics (spatial data analysis, modelling and fuzzy set algebra) have meant that current methods of land evaluation would be unrecognizable to the early practitioners. Nevertheless, there is still a need to map areas of land ('land units') and to obtain field-derived data. The quality of existing, mapped soil data (and other land attributes) in many developing countries necessitates field-based surveys. Unfortunately there is a widespread tendency amongst GIS practitioners to consider that computer manipulation and interpolation of existing data is all that is required. There is a renewed need to educate natural land resource assessors to obtain data which are relevant to the problem in hand and not to 'make do' with – often inappropriate or inaccurate but readily available – existing data.

References

Basinski, J.J. (ed.) (1978) *Land Use on the South Coast of New South Wales: a study in methods of acquiring and using information to analyse regional open space options, Volume 1, General Report*, Melbourne: CSIRO.

Chorley, R.J. (1965) A re-evaluation of the geomorphic system of W.M. Davis, in R.J. Chorley and P. Haggett (eds) *Frontiers in Geographical Teaching*, 21–38, London: Methuen.

Christian, C.S. and Stewart, G.A. (1953) *General Report on Survey of Katherine–Darwin Region, 1946*, CSIRO Australia Land Research Series No. 1, Melbourne.

Christian, C.S. and Stewart, G.A. (1968) Methodology of integrated surveys, in *Aerial Surveys and Integrated Studies*, 233–280, Paris: UNESCO.

Cocks, K.D., Ive, J.R., Davis, J.R. and Baird, I.A. (1983) SIRO-PLAN and LUPLAN: an Australian approach to land-use planning. I. The SIRO-PLAN land-use planning method, *Environment and Planning B* 10, 331–345.

Davidson, D.A. (2002) The assessment of land resources: achievements and new challenges, *Australian Geographical Studies* 40, 109–128.

FAO (1976) *A Framework for Land Evaluation*, Soils Bulletin 32, Rome: Food and Agriculture Organization of the United Nations.

FAO (1993) *Guidelines for Land-use Planning*, FAO Development Series 1, Rome: Food and Agriculture Organization of the United Nations.

Land Resources Development Centre (1966) *Land Resources Studies, No. 1 England*, Surbiton, Surrey: Land Resources Development Centre, Ministry of Overseas Development.

Nossin, J.J. (ed.) (1977) *Proceedings ITC Symposium. Surveys for Development*, Amsterdam: Elsevier.

Rowan, J.N. (1990) *Land Systems of Victoria*, Department of Conservation and Environment, Land Conservation Council, Victoria, Australia.

Stewart, G.A. (ed.) (1968) *Land Evaluation*, Melbourne: Macmillan.

Story, R., Galloway, R.W., van de Graaff, R.H.M. and Tweedie, A.D. (1963) *General Report on the Lands of the Hunter Valley*, CSIRO Australia Land Research Series No. 8, Melbourne.

WAPC (1996) *Central Coast Regional Strategy*, Perth: Western Australian Planning Commission.

ARTHUR CONACHER

LANDSCAPE SENSITIVITY

The sensitivity of a landscape to change is the likelihood that a given change in the controls of a system will produce a sensible, recognizable and sustained response. The sensitivity is a function of the propensity for change which is measured by the size of the impulse required to initiate change.

The necessary impulse is determined by the number, type and magnitude of the barriers to change. The barriers to change are complex and include the resistance of the rocks, their structure and their resilience (strength resistance); the slope, relative relief and elevation that determine the potential and mobilized energy (morphological resistance); the distance to BASE LEVEL, sources or energy or other barriers (location resistance); the ability of the system to transmit energy, waste and water, the density of pathways, stream density and joint frequency (transmission resistance); and the strength of the linkages between components of the system and the degree of hillslope-channel coupling (structural resistance).

A system also has a varying capacity to absorb the impulses of change. There are shock absorbers, filters, void spaces and energy drains. For example a BEACH disposes of wave energy by moving particles and doing work (displacement, friction, heat) and by drainage.

Sensitivity is also determined by the inheritance or degree of influence of previous system states, process systems and landforms. Of particular interest are systems that have experienced a very efficient sediment flux regime but which after environmental change or re-specification of the controls finds itself too flat or exhausted to permit further vigorous change. Large events may have a dramatic effect on a hillslope and remove all available (weathered) material. The barrier to further change is then the time (efficiency of preparation processes) needed to 'ripen' the system.

There are two limiting types of system. Mobile, fast-responding landforms have a high sensitivity, react quickly and relax to new states with facility. If they have high 'permeability' they may act as an energy filter so that change is 'skin deep' and fluctuates about an average form (e.g. a beach profile). They may also store small impulses and accumulate them into a large impulse that may cross an internal threshold value. Such systems are often morphologically complex and the correlative deposits are fragmentarily preserved. They are usually capable of rapid restoration.

Slow-responding systems may be insensitive because they are too flat, too far from boundary (energy) changes, propagate events slowly, have large storage or low concentrations of sediment transport or progressive weakening axes. Change does not take place easily so that, once their form is established, they may require large or effective events to achieve adjustment. However, if change does take place the results may be dramatic. For example, gully incision into a plain, or landsliding following progressive softening. Generally there is a persistence of relief and pattern, stagnancy of development, a palimpsest of forms and 'traditional' development in which the inherited landforms continue to develop as before even though the controlling environment may have changed.

Further reading

Begin, Z.B. and Schumm, S.A. (1984) Gradational thresholds and landform singularity: significance for Quaternary studies, *Quaternary Research* 21, 267–274.

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

——(2001) A critical assessment of the sensitivity concept in geomorphology, *Catena*, 42, 99–123.

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

Chorley, R.J., Schumm, S.A. and Sugden, D.E. (1984) *Geomorphology*, London: Methuen.

Schumm, S.A. (1985) Explanation and extrapolation in geomorphology: seven reasons for geologic uncertainty, *Transactions Japanese Geomorphological Union* 6-1, 1–18.

Thomas, D.S.G. and Allison, R.J. (1993) *Landscape Sensitivity*, Chichester: Wiley.

Thomas, M.F. (1994) *Geomorphology in the Tropics: A Study of Weathering and Denudation in Low Latitudes*, Chichester: Wiley.

——(2001) Landscape sensitivity in time and space, *Catena*, 42, 83–99. (Also see entire Volume 42 of *Catena* 2001.)

DENYS BRUNSDEN

LANDSLIDE

Landslides belong to a group of geomorphological processes referred to as MASS MOVEMENT. Mass movement involves the outward or downward movement of a mass of slope forming material, under the influence of gravity. Although water and ice may influence this process, these substances do not act as primary transportational agents. Landslides are discrete mass movement features and are distinguishable from other forms of mass movement by the presence of distinct boundaries and rates of movement perceptibly higher than any movement experienced on the adjoining slopes. Thus this group of processes includes falls, topples, slides, lateral spreads, flow and complex movements as classified by Cruden and Varnes (1996). Widespread diffuse forms of mass movement such as creep, subsidence, rebound and sagging are generally not treated as landslides.

The criteria used to distinguish different types of landslide generally include: movement mechanism (e.g. slide, flow), nature of the slope material involved (rock, debris, earth), form of the surface of rupture (curved or planar), degree of disruption of the displaced mass, and rate of movement (see MASS MOVEMENT).

Form and behaviour

There are several morphological features that can be recognized to a greater or lesser extent in most landslides (Figure 101). The uppermost part is the depletion or concave zone (erosional, generating or failure zone) where slope material has failed and become displaced downslope. In some cases the displacement may be only a few metres while in others the failure zone will be completely evacuated to expose the surface of rupture and to leave a distinctive scar on the hillslope (Plate 70). The displaced mass may remain close to the failure zone or it may continue to travel downslope leaving a transport track ending in a colluvial accumulation zone or acting as a supply to some other geomorphic agent (e.g. river, sea or glacier). The distance that landslide material travels (runout) is a characteristic of the type of landslide. For example, controlled by the height of fall and volume, rock avalanches can travel at high velocity for several kilometres. Runout distance and velocity for other types of landslide are controlled by factors such as volume, slope angle and morphology, clay content, water content, and surface frictional characteristics of the runout pathway.

Landslide movement may be instantaneous, with failure, transport and deposition taking place in a matter of seconds or minutes. Other landslides are known to have been intermittently active over tens of thousands of years, undergoing successive periods of reactivation. Eight states of activity are recognized (Cruden and Varnes 1996) including such categories as 'active' (currently moving) 'dormant' (no movement in the last twelve months but capable of being reactivated) and 'relict' (unlikely to be reactivated under present climatic or geomorphologial conditions).

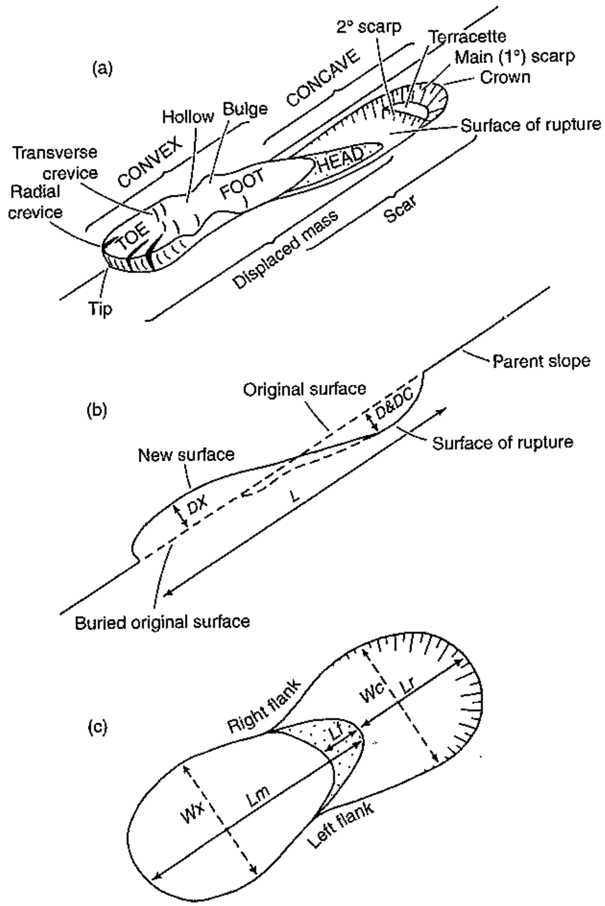

Figure 101 Morphological components of a landslide, indicating depth (D), length (L), and width (W) measurements

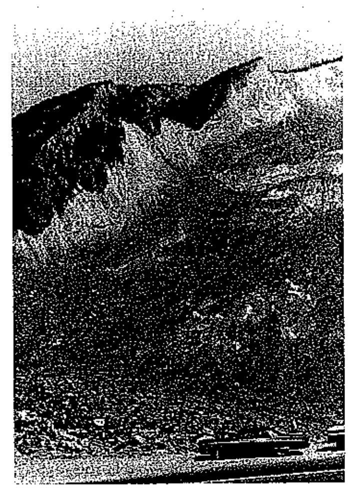

Plate 70 Hope rock slide (British Columbia), triggered by an earthquake in 1965 showing the depletion zone almost entirely evacuated of sediment

Rates of movement for different types of landslide are highly variable. Some landslides record only a few centimetres of movement a year, sustaining this rate for decades. Certain debris flows have recorded velocities of $100\,\mathrm{km h^{-1}}$ while large rock avalanches are capable of reaching velocities of $350\,\mathrm{km h^{-1}}$. Landslide deposits range in texture from dislodged blocks of intact source material to highly comminuted sediments forming a poorly sorted, unstratified diamictite.

Significance

The potential impact of landslides depends not only on velocity but also on volume. One of the largest landslides described is Green Lake Landslide, Fiordland, New Zealand which is estimated to be $13\,\mathrm{km^3}$ in volume. Submarine landslides of a similar magnitude are evident on edges of the continental shelf while recorded events such as the 1989

Ok Tedi mine landslide, Papua New Guinea and the 1970 Huascaran rock avalanche, were both estimated to exceed 50 million $\mathrm{m^3}$ in volume. Large volume, high velocity movements can also create substantial LANDSLIDE DAMs impacting fluvial systems and often posing a significant dam burst hazard.

Landslides are a manifestation of slope instability (see SLOPE STABILITY) and occur when changing conditions on a part of the slope allow shear stress to exceed shear strength. This can be brought about by a decrease in the effectiveness of factors that promote strength (e.g. a reduction in friction caused by increased pore-water pressures) or by an increase in shear stress (e.g. slope steepening). However, when a landslide takes place it changes the relative stability conditions from an unstable to a stable state, by reducing slope angle, height or weight, or by removing susceptible material. If boundary conditions are stable over a long period of geomorphic time, continued landsliding in a particular region may so alter slope conditions that other slower acting processes such as soil creep become dominant.

As most slopes are stable most of the time, landslides when they occur can be seen as a rapid and effective geomorphic response to the appearance of destabilizing changes in boundary conditions, enabling rapid adjustment and an eventual return (or tendency) toward more persistent landscape forms.

Destabilizing conditions (preparatory factors) in natural systems may be instigated by disturbances such as tectonic uplift, oversteepening of slopes by erosional activity, climatic change, deforestation or slope disturbance by human activity. The effectiveness of preparatory factors in reducing stability depends on preconditions such as material properties and slope geometry. The degree of stability afforded by preconditions and preparatory factors defines 'landslide susceptibility'. The occurrence of landslides in space can be related to susceptibility thresholds (e.g. minimum critical slope angle) while occurrence in time can be related to exceeding magnitude thresholds for triggering agents (e.g. rainfall intensity, or seismic shaking).

Landslides represent a significant hazard to life, livelihood, property, infrastructure and resources in many parts of the world. Some individual catastrophic failures have been associated with high death tolls; the 1970 Huascaran rock avalanche in Peru killed 18,000 and deaths in the 1920

Kansu landslide in China are estimated at between 100,000 and 200,000. However, much landslide damage is less obvious, seriously depleting soil resources, reducing primary productivity and destroying property by slow chronic movement.

Reference

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*, Special Report 247, Transportation Research Board, National Research Council, 36–75, Washington, DC: National Academy Press.

Further reading

Brabb, E.E. and Harrod, B.L. (eds) (1989) *Landslides: Extent and Economic Significance*, Rotterdam: Balkema.
Bromhead, E., Dixon, N. and Ibsen, M.-L. (eds) (2000) *Landslides in Research, Theory and Practice*, three volumes, London: Thomas Telford.
Crozier, M.J. (1989) *Landslides: Causes, Consequences, and Environment*, London: Routledge.
——(1999) Landslides, in M. Pacione (ed.) *Applied Geography: Principles and Practice*, 83–94, London: Routledge.
Dikau, R., Brunsden, D., Schrott, L. and Ibsen, M.-L. (1996) *Landslide Recognition: Identification, Movement and Causes*, Chichester: Wiley.
Selby, M.J. (1993) *Hillslope Materials and Processes*, Oxford: Oxford University Press.

MICHAEL J. CROZIER

LANDSLIDE DAM

Landslide dams are naturally occurring stream blockages caused by hillslope-derived MASS MOVEMENT. They represent end members in the spectrum of landforms of geomorphic HILLSLOPE-CHANNEL COUPLING, and arise from a temporary or permanent transport-limitation of the fluvial system due to excess lateral sediment delivery. The impounded natural reservoir is referred to as a *landslide-dammed lake*, whereas the term *landslide pond* relates to small water bodies perched on top of LANDSLIDE deposits, which need not necessarily be associated with stream blockage.

Landslide dams commonly form in steep terrain of many upland areas throughout the world, and constitute some of the highest natural dams on Earth. Yet the majority of occurrences are short-lived: 85 per cent of 185 worldwide examples have failed within less than one year, and nearly half within ten days (Costa and Schuster 1988). Typical failure mechanisms comprise overtopping by either lake-level rise or landslide-induced displacement waves, breaching, piping, gravity collapse of the dam face, or artificial spillway control. Rainfall, snowmelt and earthquakes are amongst the main triggers of landslides causing temporary or permanent blockage of rivers. Other mechanisms include volcanic eruptions, fluvial undercutting or, in some instances, anthropogenic activity. Stream blockages formed by LAHARS or pyroclastic flows (see PYROCLASTIC FLOW DEPOSIT) may be regarded as phenomena on a continuum between landslide and volcanic dams.

Landslide dams constitute a variety of GEOMORPHOLOGICAL HAZARDs, which may extend for considerable distances up- and downstream of the initial point of blockage. Long-term impoundment of river channels may cause extensive backwater flooding and associated SEDIMENTATION. Substantial physical impact may be created by sudden dam failures leading to catastrophic OUTBURST FLOODs of landslide-dammed lakes, which can turn into DEBRIS FLOWs, given sufficient amount of valley floor deposits to allow for alluvial bulking. Downstream reaches usually experience massive AGGRADATION, lateral channel instability and AVULSIONs, in the wake of landslide-dam failures. Rapid drawdown of the draining lake reservoir may cause further secondary loss of SLOPE STABILITY.

Landslide dams often create profound geomorphic legacies in the context of long-term landscape evolution of VALLEY floors. These include partly BURIED VALLEYs, drainage reversal, large intramontane alluvial flats resulting from the infill of landslide-dammed lakes, LAKE terraces, spillway gorges (see GORGE AND RAVINE) and RAPIDS, as well as conspicuous step-wise disruption of river long profiles (see LONG PROFILE, RIVER).

The 60-km long Lake Sarez, Tajikistan, is recognized to have been impounded by the world's highest existing landslide dam (~700 m), which has been formed by an earthquake-triggered rock avalanche of $\sim 2 \times 10^9\,m^3$ in volume near the former village of Usoi, in 1911 (Alford *et al.* 2000).

References

Alford, D., Cunha, S.F. and Ives, J.D. (2000) Lake Sarez, Pamir Mountains, Tajikistan: mountain hazards and development assistance, *Mountain Research and Development* 20, 20–23.
Costa, J. and Schuster, R.L. (1988) The formation and failure of natural dams, *Geological Society of America Bulletin* 100, 1,054–1,068.

Further reading

Korup, O. (2002) Recent research on landslide dams – a literature review with special attention to New Zealand, *Progress in Physical Geography* 26, 206–235.

SEE ALSO: dam

OLIVER KORUP

LARGE WOODY DEBRIS

Large Woody Debris (LWD) refers to fallen trees and detached wood within streams or along stream banks. The term 'large' applies to wood that is big enough to affect stream hydraulics, sediment transport, bank erosion and the resultant stream morphology, or is big enough to provide cover and habitat for aquatic organisms. In practice, LWD often refers to wood that is at least 1–2 m in length and a minimum of 10–30 cm in diameter, although the definition varies among geomorphologists and with stream size. Living trees along stream banks and vegetation outside stream channels are not referred to as LWD, although these types of wood also affect geomorphic processes (see EROSION; FOREST GEOMORPHOLOGY; RIPARIAN GEOMORPHOLOGY).

LWD is widely recognized as a critical component of aquatic ecosystems (Maser *et al.* 1988), although as recently as the 1970s management agencies removed wood from streams to reduce flooding and promote fish passage. LWD provides organic material, traps sediments, creates pools and, in general, increases the habitat diversity of streams. Resource managers now regularly emplace LWD in rivers to promote natural stream processes and recovery.

LWD enters the stream channel as logs, branches and root balls derived from bank failures, windthrow, landslides, debris flows, natural death and breakage, and human disturbances such as logging. LWD in small headwater streams often bridges the channel, only affecting morphology where breakage and decay create local log steps that promote sediment accumulation and channel widening (Nakamura and Swanson 1993). As stream size increases, the LWD falls into the stream rather than spanning it. These intermediate size streams cannot easily transport the wood, which is trapped and accumulates as single pieces or in piles known as 'jams'. The trapped wood forms log steps and pools and increases sediment storage, although channel widening can occur where logs deflect flow against a bank. In larger order streams, channel dimensions significantly exceed the size of LWD, and the wood can be transported by the flow. Larger streams sweep up the LWD as single pieces or jams along banks, on bars, and in backwater eddies (Marcus *et al.* 2002). When deposited along banks, wood in these larger systems can reduce bank erosion and promote pool development within the channel. When deposited in mid-channel, however, jams in larger streams can deflect the current and promote development of secondary channels that widen the overall stream width and increase bank erosion.

References

Marcus, W.A., Marston, R.A., Covard, C.R. Jr, and Gray, R.D. (2002) Mapping the spatial and temporal distribution of woody debris in streams of the Greater Yellowstone Ecosystem, U.S.A., *Geomorphology* 44, 323–335.
Maser, C., Tarrant, R.F., Trappe, J.M. and Franklin, J.F. (eds) (1988) *From the Forest to the Sea: A Story of Fallen Trees*, Pacific Northwest Research Station PNW-GTR-229, Portland, OR: USDA Forest Service.
Nakamura, F. and Swanson, F.J. (1993) Effects of coarse woody debris on morphology and sediment storage of a mountain stream system in Western Oregon, *Earth Surface Processes and Landforms* 18, 43–61.

SEE ALSO: debris flow; hillslope-channel coupling; landslide

W. ANDREW MARCUS

LAVA LANDFORM

Volcanic eruptions produce two main kinds of material – lava and pyroclasts (see VOLCANO). While most eruptive episodes will yield quantities of each, they can, in general, be described as predominantly effusive (producing lavas) or explosive (pyroclastic). The focus here is on the effusive variety and on extrusive rather than intrusive landforms. Lava is always partially molten on eruption but can contain a considerable volume fraction of both crystalline phases (minerals) and gas bubbles (vesicles). First-order controls on the eventual geomorphology of lava flows are provided by eruption rate and total amount of lava emitted, viscosity, the topography over which the lava flows, and the external environment (i.e. atmosphere, water or ice). Viscosity, the ratio of shear stress to strain rate, is a measure of the internal resistance to flow when a force is applied to a fluid. Magma viscosities are highly variable because there are many controlling factors, including temperature and composition of the melt. The presence of crystals increases viscosity. Bubbles, on the other hand, may increase or decrease viscosity depending on their size, properties and flow rate of the magma. Dissolved water also plays a role because it interrupts silicon bonding in the melt, hence reducing viscosity. In short, magma viscosities can vary by many orders of magnitude due to cooling, crystallization, vesiculation or loss of gas, and thereby strongly influence the nature of lava landforms.

Lava lakes and flows

Low silica, mafic lavas such as those erupted at Kīlauea (Hawai'i) have typical viscosities on eruption of only 10^2–10^3 Pa s, not much more than that of mayonnaise. When low viscosity lava erupts within a crater, it will tend to be confined within the vent region and crater, forming a lava pond or lake. Active lava lakes are connected to a deeper reservoir of magma, while passive lava lakes are not. Long-lived lava lakes are a comparatively rare phenomenon on Earth. They have been reported at several volcanoes including Kīlauea, Nyiragongo (Congo), Erebus (Antarctica) and Erta 'Ale (Ethiopia) but very few individual examples have reportedly persisted for more than a few decades. These include Halema'uma'u (Hawai'i, 1823–1924) and Erta 'Ale (probably active for at least the past century; Plate 71). Lava lakes have also been observed at oceanic ridges (Fouquet *et al.* 1995), and on the Jovian moon, Io (McEwen *et al.* 2000).

While all the manifestations discussed here are flows of lava and hence lava flows, the term 'lava flow' usually refers to erupting lava that has the opportunity to run down the flanks of a volcano or cross open ground. The expression is used

Plate 71 Erta 'Ale lava lake in the Danakil Depression (Ethiopia), is the longest-lived active lava lake on Earth. Its dimensions are approximately 80 × 90 m, and its power output is around 100–200 MW (Oppenheimer and Yirgu 2002)

both for active flows during emplacement, and for the resulting landform. In the course of a long-lived eruption, a lava flow field may develop by the superposition of many individual flow units. The current eruption of Kīlauea has yielded over 2 km^3 of lava and built up a flow field around 100 km^2 in area since it began in 1983 (a mean eruption intensity of around 3 m^3 s^{-1}, Plate 72). Areas of higher relief within a flow field that are above the 'high-tide' mark of the erupting lavas are called kīpuka. These 'islands' may preserve mature vegetation destroyed elsewhere by the active flows. Where lava enters the sea, as at Hawai'i, a lava delta or bench may build seawards, though these are often unstable features. As active lava interacts with seawater on the shore in front of the lava bench, steam explosions can construct ephemeral littoral cones.

Gas-rich, low-viscosity magma can erupt in quite spectacular fashion, with lava fountains ('fire fountains') playing to heights of up to several hundred metres. These can occur at individual vents or along fissures, which may become delineated by spatter cones or ramparts. Intense lava fountains can feed clastogenic lava flows when the spatter expelled loses little heat during its transit through the air.

The largest lava eruption of the past millennium was that of Laki (Iceland) in 1783/4.

Plate 72 The current eruption of Kīlauea (Hawai'i) has been in progress since 1983 and has formed a lava flow field some 100 km^2 in area. Most of the lava seen here is of the pāhoehoe variety, though the dark patch on the slope in the background consists of 'a'ā

It emplaced an estimated 14.7 km^3 of mafic lava, 40 per cent of which was erupted in the first twelve days at peak rates exceeding 5,000 m^3 s^{-1} (Thordarson and Self 1993). Bigger still, however, are the flood basalt eruptions or 'large igneous provinces' that punctuate the geological record. These have occurred both on land and in the oceans, and can contain 10^5–10^7 km^3 or more of lava, covering areas of 10^6 km^2, and erupted over timescales of perhaps 1 Myr, e.g. the 65-Myr-old Deccan Traps of India. The resulting landforms are known as lava plateaux. Where they are dissected by erosion, a staircase topography results from differential weathering of the rubbly boundaries and massive inner core of each superposed flow unit. This is the origin of the term 'trap', after an old Swedish word for staircase. Cooling and contraction of the cores of lava bodies can result in columnar jointing, seen spectacularly at the Giant's Causeway (Northern Ireland) and Devil's Postpile (USA).

Active lava flows radiate prodigious quantities of heat near the vent such that they rapidly form a surface crust. This can thicken sufficiently to insulate the core of the flow from thermal losses, lowering the rate of viscosity increase, and thereby promoting longer travel distance. Mafic flows quite often crust over completely, with lava continuing to flow in tunnels, which can grow in cross section by thermal erosion of the walls. When the supply of lava at the vent ceases, the last slug of lava may drain downslope leaving an empty conduit or lava tube. On Kīlauea, much of the lava flow between the Pu'u 'O'o vent, and the coastline where lava pours into the sea (a distance of over 10 km), takes place in a tunnel network, with only sporadic breakouts at the surface.

As silica content and crystallinity of lava increases, and eruption temperature drops, viscosities climb many orders of magnitude. As a result, much thicker accumulations of lava are required to overcome the resistance to flow.

Lava domes and coulées

The most silicic lavas attain viscosities of up to 10^8–10^{10} Pa s. On eruption, viscous flow is strongly resisted and slip may become concentrated along shear planes. Mounds of rock that accumulate around the vent, are called lava domes, and if they show some flow away from the vent, they are termed coulées. Sometimes, lava domes are intruded just beneath the surface causing bulges known as cryptodomes. In the weeks and days before its major 1980 eruption, a cryptodome of around 100 m in height developed on the upper flanks of Mount St Helens (USA). The eruption of Soufrière Hills Volcano (Montserrat), which began in 1995, produced an intermediate composition dome exceeding 10^8 m^3 in volume (Druitt and Kokelaar 2002). Such domes are really composite features, the construct of many individual lobes extruded at the surface of the dome (exogenous growth) and also intruded within it (endogenous growth). The frequent gravitational collapses and explosions that are typical of many dome-building eruptions can represent a severe hazard at volcanoes in populated areas. Frequent block-and-ash flows generated as the result of failure of hot sections of the dome of Merapi volcano (Indonesia) have claimed many lives.

Lava flow surface textures

The spectrum of surface textures of lava flows is impressive, and identical lava compositions can display very different textures due to subtle variations in effusion rate or cooling history. 'A'ā lava is characterized by loose clinker-like rubble, often overlying a more massive core. 'A'ā flows often build longitudinal levees by accretion along their margins. These are sometimes breached by new slugs of lava moving down the channel, and some civil protection efforts have attempted to divert flows by purposefully excavating flow levées. A curious feature of 'a'ā flows are accretionary lava balls. They form in snowball fashion as a solid core rolls along the surface of a lava flow, picking up sticky molten rock. They can reach several metres in diameter, and gain sufficient momentum to roll some distance ahead of the lava flow front.

In contrast to 'a'ā, pāhoehoe lava has a comparatively smooth surface composed of interlocking lobes, often adorned with a millimetre–centimetre-scale texture of interwoven threads. It can appear like coils of cord in ropy lava, like intestines in entrail lava, platy in slab lava, or blistered in shelly lava. Toothpaste lava forms by squeezing through cracks in the solid crust of a flow. Pāhoehoe can transform into 'a'ā lava as it moves downslope due to changes in viscosity or strain rate but the reverse transformation is never observed. Investigations of the shapes of lava flow margins have revealed further insights into pāhoehoe and 'a'ā flows. The former can be described by a scale-independent, fractal dimension, suggesting that pāhoehoe spreading is controlled by the ratio of the finely balanced forces driving advance and of those imposed by formation of the crust (Bruno *et al.* 1992). In contrast, the spread of 'a'ā flows is dominated by the driving forces, and, as a result, the shape of their margins is not fractal. These observations have potential application in interpretation of volcanic terrains on other planetary bodies.

More viscous flows often develop a blocky surface texture, consisting of fractured chunks of lava, up to several metres across, with angular facets. Giant pressure ridges called ogives sometimes wrinkle the surface of very viscous flows. Lava domes often develop spines when they are active but these are usually rather ephemeral. After the devastating 1902 eruption of Mont Pelée (Martinique), an exceptional 300-m high spine was extruded.

Smaller features surrounding openings on mafic lava flows include hornitos (literally 'little ovens'), which are chimneys or pinnacles of lava spatter and dribbles squeezed up through openings in the roof of lava tubes. Pāhoehoe flows on low slopes often display elliptical, domed structures known as tumuli. These form when the magma pressure within an active lava tunnel ruptures the overlying surface of the flow field. Fractures usually extend along the length of a tumulus, and often lava squeezes out through these and other cracks on the sides to build a larger feature.

Lavas erupted subaqueously resemble toothpaste squeezed out of a tube. The bulbous lobes that form are known as pillow lavas. These are commonplace along oceanic ridges but can also form in shallow water found in coastal, river and lake environments.

References

Bruno, B.C., Taylor, G.J., Rowland, S.K., Lucey, P.G. and Self, S. (1992) Lava flows are fractals, *Geophysical Research Letters* 19, 305–308.

Druitt, T.H. and Kokelaar, P. (eds) (2002) *The Eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999*, London: Geological Society.

Fouquet, Y., Ondreas, H., Charlou, J.L., Donval, J.P., Radfordknoery, J., Costa, I., Lourenco, N. and Tivey, M.K. (1995) Atlantic lava lakes and hot vents, *Nature* 377, 201.

McEwen, A.S., Belton, M.J.S., Breneman, H.H., Fagents, S.A., Geissler, P., Greeley, R., *et al.* (2000) Galileo at Io: results from high-resolution imaging, *Science* 288, 1,193–1,198.

Oppenheimer, C. and Yirgu, G. (2002) Thermal imaging of an active lava lake: Erta 'Ale volcano, Ethiopia, *International Journal of Remote Sensing* 23, 4,777–4,782.

Thordarson, T. and Self, S. (1993) The Laki (Skaftár Fires) and Grímsvötn eruptions in 1783–1785 *Bulletin of Volcanology* 55, 233–263.

Further reading

Francis, P. and Oppenheimer, C. (2004) *Volcanoes*, Oxford: Oxford University Press.

Hawaiian Volcano Observatory, http://wwwhvo.wr.usgs.gov/

Kilburn, C.R.J. and Luongo, G. (eds) (1993) *Active Lavas: Monitoring and Modelling*, London: University College London Press.

Rhodes, J.M. and Lockwood, J.P. (1995) *Mauna Loa Revealed: Structure, Composition, History, and Hazards*, Washington, DC: American Geophysical Union.

Sigurdsson, H., Houghton, B.F., McNutt, S.R., Rymer, H. and Stix, J. (eds) (2000) *Encyclopedia of Volcanoes*, San Diego: Academic Press.

SEE ALSO: volcano

CLIVE OPPENHEIMER

LAWS, GEOMORPHOLOGICAL

The whole concept of a scientific 'law' is far from straightforward and may be seen to become more and more problematic the further we get from the basic physical principles which can be translated into rational equations. Harvey (1969) concluded that all laws are statements of some universality, embedded within a theoretical structure. Haines-Young and Petch (1986: 13) consider that laws 'describe some characteristic or behaviour of all the members of a class of things' but that that something 'would not usually be employed in recognizing it'. They go on (pp. 14–17) to give two, contrasting examples of scientific laws: that described by Coriolis force, governing the deflection of moving bodies to right or left in the northern and southern hemispheres; and Playfair's 'Law of accordant junctions', which states that river valleys meet 'on neither too high or too low a level' (1802: 102). Whereas the first statement is a purely physical construct capable of mathematical formulation, the second is more problematic in that it is not evident how far it was intended to apply to the confluences of river channels or to valley floors (Haines-Young and Petch 1986: 16–17). Nor did Playfair formulate his observation as a law.

Since geomorphology is best seen as a historical science, it has provided many examples of both the 'qualitative' laws such as Playfair's and the 'physical' laws, such as that of Coriolis force. Although A.N. Strahler in his very influential 1952 paper called for geomorphologists to strive for 'the deduction of general mathematical MODELS to serve as quantitative natural laws' (p. 937), there has been rather limited success in that direction, since not all the phenomena of interest to geomorphology have proved equally amenable to mathematical formulation. This is recognized by Rhoads and Thorn (1996: 123) who conclude that:

> [T]he study of complex phenomena poses a problem both for inductively establishing the existence of underlying causal laws by combining mathematical formulations of these laws in predictive models. This situation may account for the fact that geomorphology has not been very successful at developing its own laws, or in using simple models that combine a few basic physical laws to predict the form and dynamics of specific landforms.

They are clearly following Strahler's exhortation in their evaluation of success or failure. Yet this is to ignore some of the most fundamental principles (or laws) which have governed the historical science of the study of landforms since the eighteenth century. We can make a crude distinction between the qualitative expressions of universality made by – most notably – James Hutton, John Playfair, G.K. Gilbert, W.M. Davis and W. Penck; and the quantitative formulations of R.E. Horton, S.A. Schumm, M.A. Melton, J.T. Hack and R.L. Shreve; and J.W. Glen.

By far the most basic proposition accepted by geomorphologists is that first recognized by Hutton in 1785, namely, that there is no need for recourse to extraordinary processes to account for the fashioning of the Earth's surface, providing we accept an almost unbelievably long lapse of time. Hutton's 'Principle of Uniformitarianism' has been much debated and travestied and – as Shea (1982) has forcefully reminded us – it remains a principle to be followed by the geomorphologist, not a 'truth' about the workings of the planet and its processes. Nevertheless, most of us would date the origin of geomorphology to Hutton and, possibly more convincingly, to John Playfair's (1802) extension and elaboration of Huttonian views. It is in the course of that development of an effectively modern geomorphological argument that Playfair described the system of valleys and 'the nice adjustment of their declivities' that came to be known as 'Playfair's Law'. In so far as fluvial valleys do, indeed, tend to show that mutual adjustment, whereas glacially eroded troughs do not, Playfair's Law is, indeed, both universal and causal.

The next significant and widespread development of principles and laws came from the United States in the latter nineteenth century: J.W. Powell's principle of BASE LEVEL; W.M. Davis's CYCLE OF EROSION and the 'laws' of the explanatory significance of 'structure, process and stage'. But there is also the whole set of propositions, explicitly termed laws, which G.K. Gilbert propounded from his great study of the badlands on the rim of the Henry Mountains, Utah, in 1877. Chorley *et al.* (1964: 550–567) discuss these contributions in some detail, which include: the law of uniform slopes; the law of declivities; the law of structures; and the law of divides. All these are seen as universal ideals, towards which a fluvial landscape will tend, but which it will never attain. One may consider the models of SLOPE EVOLUTION of W. Penck to belong in a similar tradition.

By the middle of the twentieth century, the qualitative laws discussed above were felt to be insufficiently precise. R.E. Horton (1945) set out what became known as HORTON'S LAWS of drainage basin composition (which were expanded by S.A. Schumm and M.A. Melton) covering the regularities in stream number, length, area, frequency and slope as basin order changed. To these was later added Hack's Law (Hack 1957), which relates the length of the longest stream in a drainage basin to the area of the basin (see Rodriguez-Iturbe and Rinaldo 1997). Still later, R.L. Shreve (1966) demonstrated that drainage networks could be considered to follow the statistical law of topological randomness. The most fundamental of these laws are, in some sense, related to the universal properties of networks and it is not altogether evident how far they may be related to basin geomorphology. The law which describes the deformation of ice over time – discovered by J.W. Glen in 1955 – may be more closely linked to geomorphological processes as it successfully predicts the extreme sensitivity of secondary creep in ice to changes in shear stress (Paterson 1981).

There would seem some prospect that the development of approaches such as those examined by Rodriguez-Iturbe and Rinaldo (1997) that are based upon advanced computer modelling and the concept of FRACTALS may, in time, be able to provide quantitative expressions of some of the most basic qualitative geomorphological laws (cf. Rodriguez-Iturbe and Rinaldo 1997, Chapter 6). Until that time, it must be accepted that the key laws and principles of the subject remain verbal rather than mathematical. Geomorphology is still a 'consumer' rather than a creator of quantitative physical laws.

References

Chorley, R.J., Dunn, A.J. and Beckinsale, R.P. (1964) *The History of the Study of Landforms*, vol. I, London: Methuen.

Hack, J.T. (1957) Studies of longitudinal profiles in Virginia and Maryland, *US Geological Survey Professional Paper* 294-B.

Haines-Young, R. and Petch, J. (1986) *Physical Geography: Its Nature and Methods*, London: Harper and Row.

Harvey, D. (1969) *Explanation in Geography*, London: Edward Arnold.

Horton, R.E. (1945) Erosional development of streams and their drainage basins: hydro-physical approach to quantitative morphology, *Geological Society of America Bulletin* 56, 275–370.

Paterson, W.S.B. (1981) *The Physics of Glaciers*, Oxford: Pergamon.

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.

Rhoads, B.L. and Thorn, C.E. (1996) Towards a philosophy of geomorphology, in B.L. Rhoads and C.E. Thorn (eds) *The Scientific Nature of Geomorphology*, 115–143, Chichester: Wiley.

Rodriguez-Iturbe, I. and Rinaldo, A. (1997) *Fractal River Basins: Chance and Self-organization*, Cambridge: Cambridge University Press.

Shea, J.H. (1982) Twelve fallacies of uniformitarianism, *Geology* 10, 455–460.

Shreve, R.L. (1966) Statistical law of stream numbers, *Journal of Geology* 74, 17–37.

Strahler, A.N. (1952) Dynamic basis of geomorphology, *Geological Society of America Bulletin* 63, 923–938.

SEE ALSO: base level; complexity in geomorphology; confluence, channel and river junction; Cycle of Erosion; fractal; glacier; Horton's Laws; slope, evolution; unequal slopes, law of

BARBARA A. KENNEDY

LEACHING

Leaching is the removal in solution of constituents by water or other percolating solutions. It is usually applied to chemical DISSOLUTION and removal as a liquid moves through a porous solid such as soil or a rock mass. Leaching is also used in a broader context to apply to the removal of soluble compounds from solid waste, or the extraction of metals or salts from ores. The most important factor influencing ion mobility is the amount of available water, and that in turn depends on a number of other factors. Leaching directly affects the Eh–pH conditions surrounding minerals and thus determines which elements will stay in solution. Other factors affecting the degree of leaching or weathering include texture, permeability, initial carbonate content, subaerial climate, depth of the wetting front, precipitation of secondary carbonates, soil moisture and temperature, plant transpiration, root mat extraction, porosity of the material, and depth to the saturated zone of ground water.

Leaching affects the alteration of rocks and sediments. As rocks and sediments are weathered they are altered in 3-dimensions beneath the land surface. Layers or zones of altered material may

form subparallel to the land surface. Leaching is the mechanism by which dissolved ions and clay are transferred from zone to zone. These layers or WEATHERING zones may be defined in various ways. They may differ physically, chemically and mineralogically from adjacent layers and have lateral extent. The vertical section through a stack of these zones is often referred to as a weathering profile. A soil profile is one type of weathering profile in which layers or horizons are designated. A descriptive terminology for geologic weathering zones has long been in use (e.g. Kay 1916; Kay and Apfel 1929; Leighton and MacClintock 1930; Frye *et al.* 1968) and provides a shorthand for field description of material below the solum (the A and B horizons of a soil). Hallberg *et al.* (1978) discuss the importance of standardizing weathering zone descriptive terminology and subdivide the terminology by material. The terminology describes Quaternary sediments in terms of their colour and the presence or absence of soluble carbonate minerals. In the mid-continent USA, a typical weathering profile in unconsolidated Quaternary sediments such as till consists of an oxidized and leached zone, an oxidized and unleached zone, and a reduced (deoxidized in loess) and unleached zone (Ruhe 1975). In weathered rock, zones may be determined based on the ratio of core stones to weathered matrix (Ruxton and Berry 1957). Examples of these types of weathering profiles are illustrated in Figure 102.

A few caveats are necessary when considering the use of weathering zone terminology. Colour is somewhat problematic as a descriptor because of the concepts associated with weathering zone terms such as deoxidized, reduced or unoxidized. They are not synonymous with reduced chemical states. Second, although abrupt contacts between weathering zones may coincide with stratigraphic breaks, the contacts do not indicate that stratigraphic boundaries are present. For example, a single till unit may have a prominent yellowish-brown oxidized weathering zone overlying a light-grey deoxidized zone. The zones give the unit an appearance of two tills but in actuality there is only one. Weathering zones have been related to various hydrologic conditions. In the absence of continued leaching, weathering zones function as closed systems in which chemical weathering ceases.

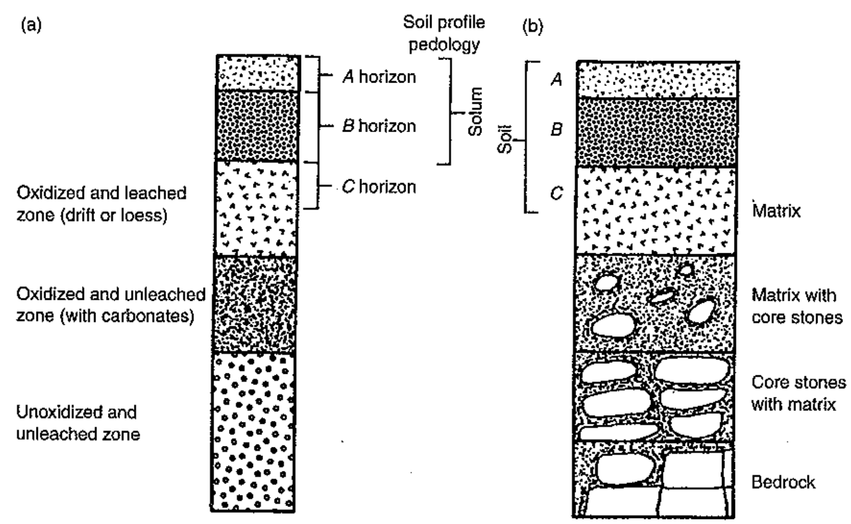

Figure 102 Weathering profiles in sediment (a) and in rock (b) with soil profiles developed in the surface materials for comparison. Modified from Ruhe (1975)

Leaching rates are important in assessing rates of groundwater contamination and plant growth. Organic chemicals are prone to be lost to leaching as their solubility in water increases. The greatest risks for hazardous leaching, rapid and unaltered dispersal away from sources, occurs in highly permeable materials such as sands that have little or no organic material. The leaching capability of introduced chemicals varies but herbicides are usually more mobile than fungicides and insecticides and should be applied with care in areas with very permeable soils.

Cover crops and no-till farming can be effective in reducing some nutrient loss through leaching. For example, the process of crop growth slows percolation and removes nitrate from solution by incorporating it into plant tissue as nitrogen. Unchecked, nitrate leaching through soils and sediments into water supplies is a serious human health and environmental contaminant.

As mentioned earlier, a common determinant for leaching is precipitation, its rate, intensity and duration. A lack of precipitation can lead to an accumulation of mobile constituents such as carbonate and other salts in soils and sediments of arid climates. Conversely, excess precipitation can lead to complete removal of these constituents from sediments in more humid environments. In artificially irrigated crop lands, a leaching requirement (LR) is usually recommended. The LR is the fraction of irrigation water that must be leached through the root zone to control soil salinity at a specific level (Foth 1984). The goal is to sustain a productive soil and produce no change in salinity during irrigation.

References

Foth, H. (1984) *Fundamentals of Soil Science*, New York: Wiley.

Frye, J.C., Glass, H.D. and Willman, H.B. (1968) Mineral zonation of Woodfordian loesses in Illinois, *Illinois State Geology Survey Circular* 427.

Hallberg, G.R., Fenton, T.E. and Miller, G.A. (1978) Standard weathering zone terminology for the description of Quaternary sediments in Iowa, in G.R. Hallberg (ed.) *Standard Procedures for Evaluation of Quaternary Materials in Iowa*, Technical Information Series 8, 75–109.

Kay, G.F. (1916) Some features of the Kansas drift in southern Iowa, *Geological Society of America Bulletin* 27, 115–119.

Kay, G.F. and Apfel, E.T. (1929) The pre-Illinoian Pleistocene geology of Iowa, *Iowa Geological Survey Annual Report* 34, 1–304.

Leighton, M.M. and MacClintock, P. (1930) Weathered zones of the drift-sheets of Illinois, *Journal of Geology* 38, 28–53.

Ruhe, R.V. (1975) *Geomorphology, Geomorphic Processes and Surficial Geology*, Boston, MA: Houghton Mifflin.

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.

CAROLYN G. OLSON

LEAST ACTION PRINCIPLE

Forms and patterns in natural systems are often the products of optimized circumstances related inherently to operational efficiency. In systems of motion, optimum operational efficiency requires, by the minimization of a system's 'action', the least expenditure of energy for completing a particular task. In other words, out of many possible alternatives nature follows the most 'economical' path. This least action principle (LAP) was formulated originally by numerous mathematical physicists, notably Pierre-Louis Moreau de Maupertuis (1698–1759), Leonhard Euler (1707–1783), Joseph Louis Lagrange (1736–1813), Sir Rowan Hamilton (1805–1865) and Carl Gustav Jacob Jacobi (1804–1851). It contains a curious and subtle twist on Newton's laws, for its variational formulation of motion does not use force and momentum but instead the physical quantities of energy and work. As a result, it often shows structural analogies between various areas of physics and has been found useful in unifying subjects and consolidating theories in various branches of science (Lanczos 1986).

By the end of the nineteenth century, LAP had become a very successful scheme applicable not only in classic mechanics, but also in electrodynamics and thermodynamics, typically through the work of Hermann von Helmholtz (1821–1894). During the twentieth century even more widespread applications were developed and in the 1910s, Albert Einstein (1879–1955) deduced the equations of *general relativity* from LAP. In the 1940s, Richard Feynman (1919–1988) identified the applicability of LAP in quantum physics and since then physicists have found that LAP also underlies the fundamental gauge theories of particle physics, leading to the establishment of what is termed *fundamental* physics (Brown and Rigden 1993). LAP has also been applied widely outside of physics. A notable example is the study of George Zipf (1902–1950), who, within the context of LAP, tried to derive the power-law form of his law for understanding the behaviour of humans on the basis of a principle of least effort (Zipf 1949). With the wide adoption of FRACTAL theory in the 1970s, Zipf's 'law' became ever popular and has been regarded as one of the essential phenomena of nature. It occurs not only in the distribution of words but also in the occurrence of cities, populations, wars, species, coastlines, floods, earthquakes and many other processes and behaviours (Schroeder 1992).

Huang and Nanson (2000, 2002) and Huang *et al.* (2002) have examined the applicability of LAP in geomorphology. Their theoretical inferences and evaluation of a wide range of case studies have shown that LAP governs fluvial systems in the form of MAXIMUM FLOW EFFICIENCY (MFE), providing a soundly based explanation as to why regular bankfull HYDRAULIC GEOMETRY relations occur in very different geographical regions. They also showed that MFE subsumes the previously proposed extremal models in geomorphology of maximum sediment transport capacity and minimum stream power. Further work is examining the application of MFE under various physical constraints (particularly available energy in the form of gradient) to explain the physical conditions for the formation of different river patterns and drainage networks.

SEE ALSO: fractal; hydraulic geometry; maximum flow efficiency; models

References

Brown, L.M. and Rigden, J.S. (1993) *'Most of the Good Stuff' – Memories of Richard Feynman*, New York: American Institute of Physics.

Huang, H.Q. and Nanson, G.C. (2000) Hydraulic geometry and maximum flow efficiency as products of the principle of least action, *Earth Surface Processes and Landforms* 25, 1–16.

Huang, H.Q. and Nanson, G.C. (2002) A stability criterion inherent in laws governing alluvial channel flow, *Earth Surface Processes and Landforms* 27, 929–944.

Huang, H.Q., Nanson, G.C. and Fagan, S.D. (2002) Hydraulic geometry of straight alluvial channels and the principle of least action, *Journal of Hydraulic Research* 40, 153–160.

Lanczos, C. (1986) *The Variational Principles of Mechanics*, New York: Dover.

Schroeder, M. (1992) *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, New York: Freeman.

Zipf, G.K. (1949) *Human Behaviour and the Principle of Least Effort*, Boston, MA: Addison-Wesley.

HE QING HUANG AND GERALD C. NANSON

LEVEE

Among the most prominent products of overbank deposition on alluvial and deltaic FLOODPLAINS and submarine abyssal plains are natural levees bordering channels. A natural levee is a wedge-shaped ridge of water-laid, channel-derived sediment which elevation tapers gently into the flanking floodbasin. Moderately to well-developed levees are present along most channel reaches, the exceptions being new, rapidly migrating, or coarse-grained braided channels.

The morphology of a levee depends upon its age, channel size and, in the case of alluvial and deltaic channels, the maximum height to which waters are ponded during floods or storm surges. Vegetation type, grain size and rate of channel alluviation exert a secondary control. Levee heights above the adjacent floodplain range from a few centimetres on young creeks to 5 m along mature sections of large rivers. Heights of submarine levees typically are greater by a factor of at least 10. Levee widths are also highly variable, ranging from fractions of a channel width to over 10 for alluvial and deltaic cases and 10 to 20 widths for submarine levees. Although data are sparse, levee cross-sectional area appears to scale linearly with channel cross-sectional area. Levee slopes transverse to the channel range widely, from virtually horizontal to near-channel values of 6° for fluvial deltaic levees, and to at most 3.5° for levees on the Amazon submarine fan. In alluvial and deltaic levees one levee of a pair is often significantly higher, wider or steeper than the other, with no systematic variation along a channel; in submarine levees of the northern hemisphere, the right-hand levee (when looking downstream) is often higher, sometimes by as much as three-halves, due to Coriolis forces.

The sedimentological characteristics of levees are highly variable. Generally it can be said that levee deposits fine laterally from coarser sands and silts closer to the channel to distal silts and clays with no vertical textural trends. Sedimentary structures and stratification consist of climbing ripple cross-laminated sands alternating with lenticular and wavy-laminated muds and rhythmically bedded, laminated, thin silts. Levees bounding submarine channels generally consist of silty to clayey spill-over turbidites with occasional thicker fine-grained sandstone beds and hemipelagic and pelagic intervals.

Levees arise by the transfer of suspended sediment from the channel to the floodplain via two mechanisms: diffusion and advection. Diffusion occurs when turbid turbulent eddies along the channel-floodplain boundary spin off onto the floodplain and decelerate, allowing grains to settle at distances determined by channel geometry, floodplain roughness, particle size distribution and flow character. Advection occurs when turbid flows leave the channel as channel water overtops the banks during the rising limbs of floods and on the outsides of meander bends.

References

Cazanacli, D. and Smith, N.D. (1998) A study of the morphology and texture of natural levees; Cumberland Marshes, Saskatchewan, Canada, *Geomorphology* 25, 43–55.

Klaucke, I., Hesse, R. and Ryan, W.B.F. (1998) Seismic stratigraphy of the Northwest Atlantic Mid-Ocean Channel: growth pattern of a mid-ocean channel-levee complex, *Marine and Petroleum Geology* 15, 575–585.

RUDY SLINGERLAND

LICHENOMETRY

Lichenometry is a dating technique using lichen measurements to supply relative or absolute dates for rock surface exposure. Geomorphologists have chiefly applied it to MORAINES and periglacial (see PERIGLACIAL GEOMORPHOLOGY) landforms in arctic and alpine environments. Other dating applications have included MASS MOVEMENTS, SCREE and blockstreams (see BLOCKFIELD AND BLOCKSTREAM), NIVATION, avalanches (see AVALANCHE, SNOW) and snow cover, LIMESTONE PAVEMENTS, seismic movements, river channel deposits, lake levels, shorelines, storm beaches and archaeological features.

The technique's dating range depends on lichen form, competition and the local environment. In temperate regions some foliose forms might survive 150 years; minutely crustose forms can provide dating over 400 to 600 years and dating may exceed 1,000 years at high latitudes. All dates can only be empirically justified unless validated by an independent source. Claims have been made for a dating range of 8,000 to 9,000 years based on radiocarbon dates for organics beneath moraines, but such a range is unlikely due to rock WEATHERING, successive glacier advances and climate change. Lichens only provide minimum dating for the last period of surface exposure while radiocarbon determinations provide maximum dates.

Three main approaches have been developed: the original approach is based on size/age correlations of largest lichens; the other two approaches are based on population size–frequency distributions, one using measurements of populations of largest lichens on uniform-aged surfaces and the other, measurements of whole lichen populations, with a population defined by its individual rock surface.

The original approach

In the original approach, pioneered in 1950 by Roland Beschel, lichen species cumulative growth rates are derived indirectly from size/age correlations of near-circular lichen bodies (thalli) growing on known-age surfaces. The longest axes of largest lichens on each surface are plotted on a graph, thallus size against date. The species cumulative growth rate is assumed to be that described by the curve traced through the largest size/age plots in the distribution. There may be a delay before species start colonizing a fresh surface; establishing this period relies on projection of the plotted growth curve on to its time axis. Absolute dates for a rock surface are obtained by fitting the largest lichen measurement to the local growth curve and relative dating may be achieved by drawing lines of equal maximum growth around areas defined by largest lichens of equal size.

Lichenometry is frequently perceived as a quick, easy, cheap method for use in areas where other dating tools are lacking. However, it has often been misapplied and results (only justifiable empirically) have been at best questionable. A review of the technique (Innes 1985) describes the difficulties surrounding species identification, particularly those in the yellow-green *Rhizocarpon* section *Rhizocarpon* group containing the species most frequently used (and confused) in lichenometric studies and the various methodologies employed, many of which were developed in attempts to resolve uncertainties arising from Beschel's five initial assumptions:

1 Thallus size correlates with age.
2 Colonization occurs soon after rock exposure.
3 Largest thalli are founder members of their populations.
4 Variations in growth rate due to habitat differences are minimized by selection of largest thalli growing in optimal conditions.
5 Accurate species identification in the field.

These are approximate assumptions because species growth rates vary over time and space, sensitively depending on habitat and climate, and identification is notoriously difficult for the non-expert. In addition, although lichens can theoretically colonize almost any bare rock surface, this may not occur until surfaces are WEATHERED. Consequently, both growth rates and delay before colonization should be confirmed at each dating site, a difficult procedure in places where there are no dated surfaces for growth-rate calibration.

Additional problems for this approach introduced during its development, arose as a result of misconceptions of the nature of lichen growth and confusions in terminology. The two terms causing the greatest confusion are 'great period' and 'lichen factor'.

Beschel (1961: 1,046, 1,057) described lichen growth as sigmoidal: beginning very slowly then gathering speed with many thalli passing through a 'great period' (limited to a few decades) before dropping to a long-term constant value. However, geomorphologists, due to a scarcity of younger thalli on surfaces of known age in their field areas, have often missed the initial part of the growth curve and assumed that the 'great period' is represented by linear growth over the whole historical period, possibly lasting some 300 years, with this followed by apparently declining growth, the decline being suggested by a curve drawn through very few lichen-size/radiocarbon-age correlations. However, as noted above, radiocarbon cannot reliably be used to date the latest period of moraine exposure and protracted growth curves constructed on this basis are therefore questionable.

These misconceptions contributed to the confusion over interpretation of the 'lichen factor' a term employed in many studies to describe the average value of maximum growth over 100 years, despite Beschel's (1961: 1,055) assertion that when calculating a 'lichen factor' the 'great period' should not be taken into consideration and that averaging cannot describe growth rates. This is because a straight line drawn from its origin and projected over 100 years will only cut a sigmoid curve in one place. Beschel's intention was that the 'lichen factor' should be used as a measure of hygrocontinentality and he defined the term as a growth velocity gradient produced by low precipitation at high altitudes that could be a helpful indicator when planning mountain reservoirs. And he warned that no standard growth velocity could be expected across a large region.

Population size-frequency approaches

In attempts to resolve some of the uncertainties in the initial assumptions and supply a measure of dating independence, two population size-frequency approaches have been developed. In the first, a composite curve is computed for the frequency distribution of large samples of longest axial measurements of largest lichens growing on uniform-age surfaces. The curve can be decomposed into subpopulations describing discrete events in their relative-age order. Where absolute dates are required, means of the largest lichens on each event surface are either plotted or regressed against dates of historical events to obtain a regional growth curve. Such a curve is, however, questionable (as Beschel warned) and requires careful testing before acceptance.

The advantages of this approach are that it can be used to investigate the history of seismic and large diachronous surfaces; statistical methods, with error limits and modelling can be applied, and anomalous lichens (either coalesced thalli or survivors from an earlier population) are less likely to corrupt the data.

In the second size-frequency approach, longest axial measurements are taken of whole lichen populations, with populations growing on surfaces of similar aspect and lithology and containing not less than thirty individuals, but ideally over 100 (statistical tests can be used to show where the smaller populations may safely be grouped). Changes in the shape of unimodal population size-frequency distributions show the relative age of populations. Bi- or multi-modal distributions indicate surface changes that have partially removed lichen thalli creating space for new colonizers; isolated large thalli may either be survivors or anomalous growths.

The advantages of this approach are similar to those of the largest lichen size-frequency one, but in this case a single boulder's whole colonization history is reflected by its population distribution, and comparison of its history with the histories of other surrounding surfaces provides insights into micro-environmental effects; showing for example, colonization rates differing on surfaces of differing aspects. Hence, colonization delays should be investigated before dating the proximal and distal sides of moraines.

The limitations of these two approaches are that a very large number of lichen measurements are required and the quality of absolute dating depends on the reliability of the independent dating source. Consequently, neither size-frequency approach achieves the desired status of full independence in absolute dating. However, because of large data sets the uncertainties implicit in the initial assumptions are less important.

Lichenometry is a reliable relative dating tool and, used with proper care, can provide revealing insights into land-forming processes especially in locations where the technique can be used within some independent dating framework supplied, for example, by historical records or DENDROCHRONOLOGY. In these circumstances absolute dating, frequencies and rates of change can be established with a reasonable degree of confidence.

References

Beschel, R.E. (1961) Dating of rock surfaces by lichen growth and its application to glaciology and physiography (lichenometry), in G.O. Raasch (ed.), *Geology of the Arctic II. 1st International Proceedings, Arctic Geology*, 1,044–1,062, Toronto: University of Toronto Press.

Innes, J.L. (1985) Lichenometry, *Progress in Physical Geography* 9, 187–254.

Further reading

Beschel, R.E. (1973) Lichens as a measure of the age of recent moraines, *Arctic and Alpine Research* 5, 303–309.

Bull, W.B., King, J., Kong, F., Moutoux, T. and Phillips, W.M. (1994) Lichen dating of coseismic landslide hazards in alpine mountains, *Geomorphology* 10, 253–264.

McCarroll, D. (1993) Modelling late-Holocene snow-avalanche activity: incorporating a new approach to lichenometry, *Earth Surface Processes and Landforms* 18, 527–539.

Winchester, V. and Harrison, S. (1994) A development of the lichenometric method applied to the dating of glacially influenced debris flows in Southern Chile, *Earth Surface Processes and Landforms* 19, 137–151.

VANESSA WINCHESTER

LIDAR

LIDAR light detection and ranging, is a form of airborne scanning altimetry which is of great importance for mapping landforms and landform change, especially at relatively local scales. It can be used to produce DIGITAL ELEVATION MODELS and provides high-resolution data on topography. It has been used in a number of geomorphological applications including cliff and landslide monitoring, the study of tidal channels, assessing subsidence risk, predicting areas subject to storm surges and tsunamis and changes in beach height (see, for example, Adams and Chandler 2002; Brock *et al.* 2002; Stockdon *et al.* 2002).

References

Adams, J.C. and Chandler, J.H. (2002) Evaluation of lidar and medium scale photogrammetry for detecting soft-cliff coastal change, *Photogrammetric Record* 17, 405–418.

Brock, J.C., Wright, C.W., Sallenger, A.H., Krabill, W.B. and Swift, R.N. (2002) Basis and methods of NASA airborne topographic mapper lidar surveys for coastal studies, *Journal of Coastal Research* 18(1), 1–13.

Stockdon, H.F., Sallenger, A.H., List, J.H. and Holman, R.A. (2002) Estimation of shoreline position and change using airborne topographic Lidar data, *Journal of Coastal Research* 18(3), 502–513.

A.S. GOUDIE

LIESEGANG RING

Liesegang rings are alternating concentric iron-rich and iron-depleted shells which are formed by differential chemical leaching and precipitation in rocks exposed to WEATHERING. The rings (shells) are developed in rock blocks, typically 2–20 cm in diameter and 10 cm thick (Liesegang blocks), bound by the tectonically induced joints at the periphery, and joint or bedding planes at the bottom. The Liesegang rings follow the configuration of the outer shape of the blocks. Structurally, the Liesegang blocks can be divided into four types: (a) primitive (single shell), (b) ordinary (multi-shell), (c) compound (multi-pattern), and honeycomb (cylindrical) Liesegang blocks. The most important factors involved in transformation of a bed of rock into Liesegang blocks are: a condensed grid of joint polygons, surface water, the appropriate topographic site of exposure, and composition, texture and thickness of the beds. Field, hand specimen, microscopic and geochemical studies suggest two opposing diffusion direction trends in the course of Liesegang ring formation: one mainly Si-trend (outward), and one mainly Fe-trend (inward).

The effect of recent tectonic movements on the Liesegang blocks manifests itself in various forms. Of interest is the partial replacement of the earlier Liesegang patterns by the later patterns along joints, and formation of compound Liesegang blocks (Liesegang blocks with more than one set of Liesegang patterns). From the study of Liesegang patterns within the compound Liesegang blocks, the configuration of the joint polygon related to each Liesegang pattern and thereby the sense of stress field migration in the area, can be deduced.

Further reading

Shahabpour, J. (1998) Liesegang blocks from sandstone beds of the Hojedk Formation, Kerman, I.R. Iran, *Geomorphology* 22, 93–106.

SEE ALSO: WEATHERING

JAMSHID SHAHABPOUR

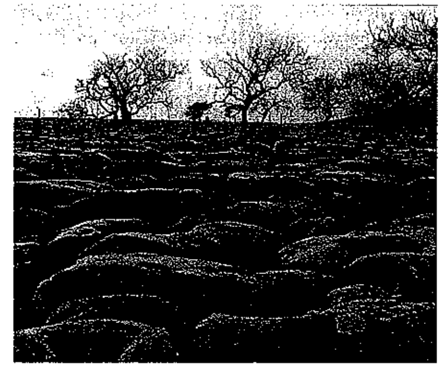

Plate 73 Stepped limestone pavement with well-developed karren features, Triglav area, Julian Alps, Slovenia

LIMESTONE PAVEMENT

Limestone pavements are stripped bedding-plane rock surfaces constituting complex polygenetic assemblages of KARREN landforms, or Karrenfeld. The limestone's bedding planes are scoured of mature weathering forms and debris producing a surface on which develop solutionally widened fissures (grikes, Kluftkarren) and residual limestone blocks (clints, Flachkarren).

Adjectives such as 'regular' or 'rectangular' popularly describe clint and grike patterns, reflecting DISSOLUTION along the two main joint directions at right angles usually found in limestone massifs. However, field situations show highly complex and variable dissection patterns. In addition, not all limestone pavement areas are on single bedding planes. They can be benched or stepped across several beds depending on the limestone outcrop (Schichttreppenkarst, sheet-stepped karst) (Plate 73, and see KARREN, Plate 67). Whilst the term 'pavement' implies a walkable surface, there is a spectrum of surface limestone bedding-plane landforms, from highly dissected, almost shattered strewn rock fields, to virtually undissected rock sheets.

Morphometric work (see GEOMORPHOMETRY) comparing numerous pavements identifies typical ranges of clint and grike dimensions (Goldie and Cox 2000). Clints up to 1 m long are commonest, but many are longer, typically up to several metres. Whilst most widths are up to 500 cm, many are over 1 metre. Most grikes are up to 50 cm wide, whilst grike depths have a wider range, between 20 and 150 cm. These ranges, however, do not encompass all limestone pavement surfaces. For example, although 95 per cent of measured grikes were less than 30 cm wide, the absolute range is from 1 cm to over 1 m; although a 'typical' grike width is 13 to 16 cm. Grike depths showed a 70-fold variation from 4 to 274 cm but focus on 100 cm. A rather weak relationship between grike widths and depths suggests that their horizontal and vertical development are not strongly linked.

Clint shape as demonstrated by a clint width to length ratio (theoretically ranging from 0 to 1) is interesting in relation to conventional ideas of pavements, as a ratio of 0.4 is four times more common than a ratio of nearly 1 (indicating square). Very elongated clints are also rare (ratios close to 0). This implies that one joint direction dominates the planform of limestone pavements. A complication, however, is that the two main joint directions in limestone massifs are frequently bisected, resulting in many triangular and diamond-shaped features, causing problems for measurement and interpretation of rectangularity.

Beyond measurability are features of extreme or minimal dissection. Where clints are too narrow or small they move easily and fall over, failing to satisfy the walkability concept. These should be excluded from the limestone pavement definition. Harder to exclude, however, is the undissected extreme where the surface lacks grikes but is certainly walkable. Although lacking the karren forms expected on limestone pavement, such bare rock expanses can really only be called pavement. The more dissected extreme shows a geomorphological transition to a rock field.

Plan morphometric data says nothing about cross-sectional form of clints and grikes, important for understanding pavement weathering sequences. Depth and intensity of surface weathering may not affect clint area but changes their vertical relief (see Figure 103). For example, clints become pinnacled or very well rounded where solutional weathering is advanced (Plate 74).

The main factors producing a good limestone pavement are strong limestone, with some fractures, but not so fractured as to weaken the beds too much. Scouring is necessary to clear weathered debris off bedding planes, and pavements then need to be in a favourable solutional environment, i.e. reasonably wet and possibly covered with soil and vegetation, to develop their characteristic karren. Then that cover needs removal.

Glaciation is the scouring mechanism creating most of the world's limestone pavements. Other scouring agents include marine stripping, scarp retreat in semi-arid conditions and even human removal. In the main, though, distribution of limestone pavements is related to glaciation and thickly bedded, strong, pure limestones. Thus there is much in the younger mountain ranges such as the Alps, Rockies and Himalayas and in shield areas, such as northern Canada. Vast expanses of pavement on dolomites are found in northern Canada.

Rock type influences pavements at several scales. The best, most persistent, pavements develop on pure, mechanically strong and massively bedded limestones and dolomites. Similar landforms can occur on sandstones but weathering and soil development here generally mean that the features do not last except in situations such as waterfall or river beds where constant fluvial action keeps jointed outcrop clear. Pure limestones stay clear as soil formation is exceptionally slow due to the lack of insoluble residue, and the

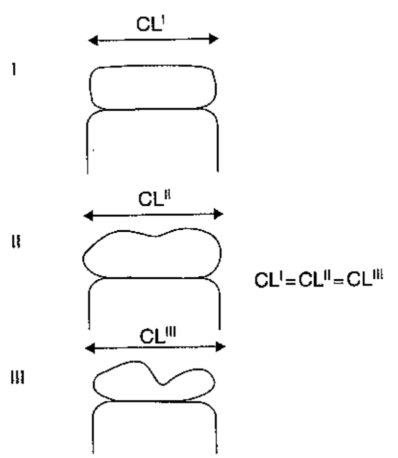

Figure 103 Simple cross-sectional diagrams of three clints demonstrating how plan measures of clint size (CL^I, etc.) can be identical on clints with very different degrees of down-wearing

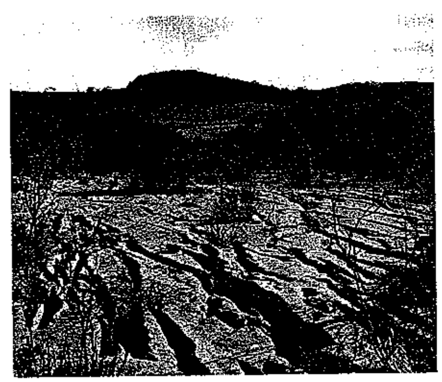

Plate 74 Limestone pavement with a slight slope (5° to 8° down away from viewer) at Gaitbarrows, Cumbria, UK. There are several stages of grike development from relatively immature short slits to grikes running for tens of metres. Other karren features are less well developed

swallowing of material by developing solutional fissures. Thin or weak limestones (internal weaknesses, either horizontal weaknesses or frequent vertical joints within a given bed) favour weathering both by mechanical and solutional means, permitting easy rock breakup unable to sustain a pavement.

Slope significantly affects both how glacial scour operates and the subsequent pavement surface development by solution (see Plate 67). Sloping layered outcrops provide complex situations for glacial scour. The postglacial surface may be interrupted by areas where shelter from the ice allowed preglacial karstified features to survive. This is found in north-west England and in the Alps, for example. Some pavements may also have characteristics surviving from earlier phases of karstification (see PALAEOKARST AND RELICT KARST).

Another factor affecting pavement form and distribution has been human activity. Some limestone pavements are anthropogeomorphic (see ANTHROPOGEOMORPHOLOGY), indirectly because their exposure results from soil erosion, and directly because of clint removal by humans for resource reasons. Because their botanic as well as landscape value is threatened by this, limestone pavements in the UK are now legally protected (Goldie 1993).

References

Goldie, H.S. (1993) The legal protection of limestone pavements in Great Britain, *Environmental Geology* 21, 160–166.

Goldie, H.S. and Cox, N.J. (2000) Comparative morphometry of limestone pavements in Switzerland, Britain and Ireland, *Zeitschrift für Geomorphologie N.F. Supplementband* 122, 85–112.

Further reading

Ford, D.C. and Williams, P.W. (1989) *Karst Geomorphology and Hydrology*, London: Unwin Hyman.

Fornos, J.-J. and Gines, A. (1996) *Karren Landforms*, Palma: Universitat de Illes Balears.

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

Vincent, P.J. (1995) Limestone pavements in the British Isles: a review, *Geographical Journal* 161(3), 265–274.

SEE ALSO: karren; karst; limestone pavement

HELEN S. GOLDIE

LINEATION

Lineations are linear patterns observed on imagery that represent fractures. The fractures may be either joints or faults (see FAULT AND FAULT SCARP). Each lineation does not necessarily represent an individual fracture, and typically represents a fracture zone. Lineations include, but are not limited to, actual joints in bedrock, straight stream segments, linear alignments of natural vegetation, aligned topographic features and linear changes in image tone or texture. Any feature thought to be a lineation should be one of a group of parallel features, and each lineation and group (set) may consist of different types of patterns. For instance, a line of vegetation may continue as a linear pattern of dark-toned soil across a light-toned field, and then as a stream valley.

Lineations are readily apparent on all scales, resolutions and types of imagery – aerial photographs, radar and various types of satellite imagery – on hard copy, as well as in digital form. The delineation of lineations is, however, both art and science (Wise 1982). For best results, and particularly if quantitative results are desired, the imagery should be vertical or near vertical. Certain features, such as those noted above, define lineations, but the interpretation of these features, particularly tonal and textural variations, can be difficult. The eye must be trained to see them, and the more experience one has, the greater the number of lineations that can be identified with confidence on any given image. All hard-copy images should be evaluated meticulously from different angles and with different kinds and angles of illumination. Some lineations will be obvious in sunshine, for example, but cannot be seen in natural light on a grey, overcast day. Topographic lineations are often more easily identified using stereo imagery, but those defined by vegetation or tonal variations can be readily – and more easily – identified monoscopically. Although attempts have been made to automatically delineate lineations on digital imagery, they have not been generally successful (e.g. Ehlen *et al.* 1995). Finally, the individual delineating the lineations should also be careful not to confuse linear man-made features, ancient or modern, with natural lineations.

Stereonet projections can be used to display lineation orientation data so that sets can be identified (dips are assumed vertical). Different patterns are indicative of fracture formation under different tectonic conditions, and the chronology of the different events can be identified using cross-cutting relations between individual lineations and lineation sets. When the lineations are in digital form, quantitative information on length and spacing (frequency) can also be obtained. These types of data are useful with respect to landform evolution and engineering geomorphology, and can provide information on fracturing in areas that are difficult to access in the field. Furthermore, lineations can be directly related to joint or fault data identified on the ground. Several studies have identified lineations on imagery, then located the individual lineations on the ground (e.g. Lattman and Parizek 1964). Other studies have shown statistical links between joints and lineations (e.g. Ehlen 1998, 2001).

Another term often used for lineation is 'lineament'. Lineaments are in fact subsets of lineations: they are topographic features and, for example, do not include linear patterns produced by tonal changes, which are among the most important indicators of fractures on imagery. Examples of lineaments include aligned saddles or ridge lines and aligned stream valleys.

References

Ehlen, J. (1998) A proposed method for characterizing fracture patterns in denied areas, in J.R. Underwood, Jr and P.L. Guth (eds) *Military Geology in War and Peace*, Boulder, CO: Geological Society of America Reviews in Engineering Geology 13, 151–163.

——(2001) Predicting fracture properties in weathered granite in denied areas, in J. Ehlen and R.S. Harmon (eds) *The Environmental Legacy of Military Operations*, Boulder, CO: Geological Society of America Reviews in Engineering Geology 14, 61–73.

Ehlen, J., Hevenor, R.A., Kemeny, J.M. and Girdner, K. (1995) Fracture recognition in digital imagery, in J.J.K. Daemen and R.A. Schultz (eds) *Rock Mechanics, Proceedings of the 35th U.S. Symposium on Rock Mechanics*, 141–146, Brookfield, VT: A.A. Balkema.

Lattman, L.H. and Parizek, R.R. (1964) Relationship between fracture traces and the occurrence of ground water in carbonate rock, *Journal of Hydrology* 2, 73–91.

Wise, D.U. (1982) Linesmanship and the practice of linear geo-art, *Geological Society of America Bulletin* 93, 889–897.

SEE ALSO: jointing; remote sensing in geomorphology

JUDY EHLEN

LIQUEFACTION

Liquefaction is the transformation of granular material from a solid to a liquid by an increase in pore pressure. Liquefaction is associated with earthquakes, with saturated ground, with granular sandy/silty soils, with some spectacular building collapses, and with submarine landslides and waste tip failures. During the 1964 earthquake at Niigata in Japan large apartment buildings were gently tilted by as much as 80° on to their sides, with no structural damage, as a result of liquefaction of a shallow sand bed underlying the structures.

In the Kobe earthquake of January 1995 widespread damage occurred. Massive soil liquefaction was observed in many reclaimed lands in Osaka Bay including two man-made islands. Both islands had been constructed from fill material derived from decomposed granite. The grain size of the fill varied from gravel and cobble-sized particles to fine sand, with a mean grain size of approximately 2 mm. In the Port Island ground, liquefaction of this fill material (the thickness of the submerged fill in this island was about 12 m) resulted in settlement of between 0.5 and 0.7 m. Soil liquefaction also occurred around the port of Kobe and caused extensive damage to many industrial and port facilities such as tanks, wharves, quay walls, cranes, and the collapse of the Nishinomiya Harbour Bridge.

A catastrophic liquefaction failure occurred in a coal waste tip at Aberfan in Wales in 1966. The tips from the Merthyr Vale Colliery were sited on the upper slopes of the valley sides, directly above Aberfan, and some 100 m up on a 13° slope. These slopes consisted of the permeable Brithdir sandstone, with many springs and seeps in evidence. Tip no. 7 was built over these springs and in October 1966 a flowslide developed in the saturated material – a classic liquefaction event. The liquefied tip material ran into the village and killed 144 people, most of whom were children in the village school.

In the Netherlands liquefaction events often concern flowslides in underwater slopes of loose sand. Flowslides are observed after sudden liquefaction of the sand under static loading. The liquefaction is caused by the contraction (packing collapse) occurring simultaneously in a large volume of sand and the impossibility of rapid drainage of the completely saturated pore fluid. The upper 10–30 m of soil in the western part of the Netherlands consists of Holocene layers of clay, peat and sand. Tidal estuaries were formed in Zeeland, the south-west part of the country, from the beginning of the Holocene. The position of the tidal estuaries, however, has shifted in an alternate process of rapid erosion and sedimentation. Rapid sedimentation has resulted in thick layers of loose sand at many locations. Median grain diameter is around 200 μm; with a silt content of perhaps 3 per cent. Liquefaction is accompanied by sudden pore pressure increase as the open sand packing collapses from a metastable state to a more stable state.

Further reading

Ishihara, K. (1993) Liquefaction and flow failure during earthquakes: 33rd Rankine lecture, *Géotechnique* 43, 351–415.

Nieuwenhuis, J.D. and de Groot, M.B. (1995) Simulation and modelling of collapsible soils, in E. Derbyshire, T. Dijkstra, and I.J. Smalley (eds). *Genesis and Properties of Collapsible Soils*, 345–360, Dordrecht: Kluwer.

Smalley, I.J. (1972) Boundary conditions for flowslides in fine-particle mine waste tips, *Transactions of the Institution of Mining and Metallurgy*, section A 81, 31–37.

Ter-Stepanian, G. (2002) Suspension force and mechanism of debris flows, *Bulletin of the International Association of Engineering Geology and the Environment* 61, 197–205.

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

SEE ALSO: quickclay

IAN SMALLEY

LITHALSA

The term lithalsa was used for the first time by S. Harris (1993) to describe mineral mounds in Yukon (Canada). They are mounds similar to PALSAS but without any cover of peat. The upheaval of the soil is the result of the formation of segregation ice in the ground. These mounds which, like the palsas, are found in the domain of discontinuous PERMAFROST, were successively called purely minerogenic palsas with no peat cover, mineral palsas, cryogenic mounds, mineral permafrost mounds and palsa-like mounds. Lithalsas are presently known only in Subarctic Northern Québec and Lapland in places where the temperature of the warmest month is between +9 and +11.5 °C and the mean annual temperature between −4 to −6 °C.

Remnants of lithalsas formed during the Younger Dryas exist in Ireland, Wales and on the Hautes Fagnes plateau in Belgium (Pissart 2000). They are closed depressions surrounded by a rampart. These features were first explained as remnants of pingos. They are now regarded as periglacial phenomena which provide more precise palaeoclimatic indications.

References

Harris, S.H. (1993) Palsa-like mounds developed in a mineral substrate, Fox Lake, Yukon Territory, in *Proceedings of the Sixth International Conference on Permafrost*, 5–9 July, Beijing, China, South China University of Technology, vol. 1, 238–243.

Pissart, A. (2000) Remnants of lithalsas of the Hautes Fagnes, Belgium: a summary of present-day knowledge. *Permafrost and periglacial processes* 11, 327–355.

ALBERT PISSART

LITHIFICATION

The process by which unconsolidated sediments become indurated sedimentary rock, originating from Greek and Latin for 'to make rock'. The source of sediments is typically loose material that accumulates upon the surface as a result of subaerial weathering. These are then transported and deposited before the process of lithification begins. Lithification precedes DIAGENESIS and must not be confused with the term (diagenesis refers to the processes and their products affecting rocks during their burial). The process of lithification involves three main components. First, desiccation (drying of sediment) reduces the amount of pore space within the sediment by eliminating the water within the material. Desiccation is particularly significant in fine-grained sediments (clays and silts), where water cannot percolate easily through the pores. Compaction further reduces pore space, aided by the increasing weight from the materials above as surface deposition continues. Surface tension between the individual grains then allows them to act as a consolidated mass.

Second, cementation binds the loose sediment together by filling in the remaining pore spaces. It is typically done through precipitation of cementing agents in solution as water flows through the pore spaces. The most common cementing agents are calcite ($CaCo_3$), dolomite ($CaMg[Co_3]_2$), quartz (SiO_2), and iron oxide (Fe_2O_3) (haematite).

The pH of a solution plays a significant role in determining the cementing agent type. An acidic solution will tend to produce a quartz cementing agent, whereas an alkaline solution will invariably generate a calcitic or dolomitic cement.

Third, crystallization is a specific form of cementation. It is particularly effective on carbonate sediments, whereby crystals form in pore spaces from minerals in solution and bond onto existing crystals within the sediment. This often has the side effect of making some rocks harder.

Further reading

Maltman, A.J. (1994) *The Geological Deformation of Sediments*, London: Chapman and Hall.

STEVE WARD

LOESS

Loess is a sedimentary deposit, largely composed of coarse and very coarse silt, which is draped over the landscape. The silt is largely quartz and, at the soil structure level, the loess particle system has an open packing, which is a result of aeolian deposition during Quaternary times. This open structure gives rise to the main practical problem; when loess ground is loaded and wetted the structure may collapse and subsidence occurs. Construction on loess has to be carefully designed in order to avoid these subsidence problems. The silt-sized particles and the open metastable structure are the main characterizing features of loess deposits.

The literature on loess is vast and complex; the major loess languages are Russian, Chinese, English, German and French. The largest literature is in Russian (see Kriger 1965, 1986; Trofimov 2001) but there are major works in many languages. Scheidig (1934) was for many years the standard work, possibly until supplanted by Pye (1987) and Rozycki (1991). Early literature has been collected (Smalley 1975) and an outline bibliography attempted (Smalley 1980). Loess investigation has had a major part in the activity of the International Union for Quaternary Research (INQUA), which publishes the one specialized journal – *Loess Letter*. The one institute in the world totally devoted to the study of loess is the Xian Laboratory for Loess and Quaternary Geology of the Chinese Academy of Sciences; there is a focus of geotechnical activity at Moscow State University and at the G.A. Mavlyanov Institute of Seismology in Tashkent.

Loess really is an aeolian material; it is the aeolian factor which gives it its defining characteristics, and separates it from other deposits. Loess is a wind-blown silt; the aeolian deposition accounts for the open metastable structure, which accounts for the propensity to collapse when loaded and wetted. The wind-blown deposition mode accounts for its characteristic geomorphological position – the 'draping' over the landscape of the loess on the ground, thinning away from source. Loess in the air is a dust cloud, travelling in low suspension over a relatively short distance. In parts of the world where the geomorphology is favourable this dust has been depositing for millions of years (but see also Pecsi 1990).

Loess formation

An area rich in controversy and argument; how do loess deposits form? Even in the twenty-first century the tag-end of this debate still carries on – but only in the Russian literature. By the end of the nineteenth century it was fairly widely accepted that the key event in the formation of a loess deposit was deposition by aeolian action. This might be called the mainline loess view; loess is an airfall sediment.

But starting early in the twentieth century and continuing fitfully throughout the entire century there has been an alternative view of loess formation, which has generated a large literature, from proponents and opponents. This can be called the 'soil' theory of loess formation, or the '*in situ*' theory. At its heart is the idea that loess is formed where it is found, by a process of 'loessification', the conversion of non-loess ground into loess ground. This is largely a Russian theory and its chief, and very forceful protagonist, was L.S. Berg. The minor role was played by R.J. Russell in the USA who put forward a very similar theory to explain the Mississippi valley loess in 1944.

Another question, which arose somewhat later in the history of loess investigation, was that of the possible mechanism for the formation of the vast amounts of quartz silt that were required to make large loess deposits. What sources of geo-energy were available for large-scale silt formation? And did this affect the nature and distribution of loess deposits? This question connected in particular to the problem of 'desert' loess.

V.A. Obruchev, early in the twentieth century, placed loess at the fringes of hot, sandy deserts. This desert loess has been something of a problem ever since. B. Butler, a noted Australian soil scientist, suggested that it did not exist, because he could not find it in Australia; and Albrecht Penck, a famous geomorphologist, suggested that peri-Saharan deposits were essentially lacking. So the question arose, could large hot sandy deserts generate enough loess-sized material to form significant loess deposits? In geographical terms it appears that the Chinese and Central Asian deserts do have major loess deposits close by, but the Sahara and Australian deserts do not. There is a loess-like material, called 'PARNA' by Butler, in south-east Australia, but it has modest extent. It has been proposed that the Central Asian and Chinese loess deposits are also near very high mountain regions, and that these are the real source of the loess particles, with the deserts simply acting as holding areas or reservoirs.

Loess distribution

There are some very large deposits, and many widespread smaller ones. The classic deposit is in China, in the north and north-west of the country; this is the 'Yellow Earth', which gives colour to the Yellow River, and played a major part in the development of the Chinese civilization – the only one of the ancient civilizations which has lasted into modern times. Deposits in Lanzhou are believed to be over 300 m, perhaps over 400 m, thick. Although loess is thought of as a Quaternary material there are suggestions that loess deposition in China has been going on since the Miocene, i.e. for over 20 million years.

The other large deposits are in central North America, relating to the Great Plains, and the Dust Bowl; in South America, the Pampas; and in Europe. The European deposits are complex but might be divided into the northern glacially related deposits, and the Danubian deposits which derive their materials from mountain sources. Of the smaller deposits those in North Africa and in New Zealand are of interest. There are deposits near the coast in Tunisia and Libya which are certainly very loess-like, but the particle size is rather large. The Sahara is a great source of wind-blown dust but most of it is very small, which travels in high suspension for great distances. In New Zealand there are significant loess deposits in the North and South Islands, associated with the mountains. New Zealand is the only country which has a monograph devoted to the national loess (Smalley and Davin 1980) and it is to be hoped that more countries will follow this lead.

With the disappearance of the Soviet Union and the revision of geography in the fringe regions new loess-rich countries have emerged, in particular Ukraine and Uzbekistan. Ukraine has large loess cover and in many parts chernozem soils have developed in the loess providing top-grade agricultural land. Also, in Kyiv is the most amazing loess building, the Pecherskaya Lavra, the Caves Monastery, where the monks have excavated into the loess and developed a subterranean complex in the Dnepr loess. In Uzbekistan, in the eastern part, near the Tien Shan mountains, are major loess deposits. The capital city of Tashkent is built on loess and was largely destroyed by a large earthquake in 1966. Loess ground is very vulnerable to earthquake shocks.

Loess stratigraphy

The basic idea of loess stratigraphy was invented by John Hardcastle in Timaru, New Zealand in 1890; this was that a loess deposit can give a good indication of past climates. Loess acted as a 'climate register'. At the beginning of the twentieth century problems of climate change were exciting quite a lot of interest and this drove a large research effort devoted to loess stratigraphy.

Hardcastle's insights were ignored; as in the case of Mendel and genetics, the scientific culture was not prepared for them. They were addressing a problem that had no structure, no framework, no reference points. It appears that stratigraphic ideas re-emerge with Soergel in Germany in 1919 and developed slowly in Europe. The 'eureka' moment came at the INQUA Congress in 1961 in Poland when Liu Tung-sheng presented the work of the Chinese investigators which showed multiple palaeosols in the Luochuan loess. This thick loess section showed alternating layers of loess and palaeosols. The palaeosols indicated warm periods and the loess layers cold periods – a climatic indicator as Hardcastle had suggested. The data from Luochuan suggested many climatic oscillations in the Quaternary period and laid the foundations for continuing investigations. It was a giant step away from the simple four event Quaternary derived from Alpine observations.

Applied geomorphology

Loess drapes the landscape; it is the accessible ground in which engineers operate. The most significant and expensive of the loess ground engineering problems is hydroconsolidation and soil structure collapse, caused by loading and wetting, and leading to subsidence and structural failure. This was classically a problem in the Soviet Union and now occurs in post-Soviet states like Ukraine and Uzbekistan. The building of irrigation canals in loess in Uzbekistan during one of the early five-year plans alerted the Soviet authorities to the vast problems of subsidence. It was, more than anywhere, a Soviet problem. Therefore most of the subsidence literature is in Russian (see Trofimov 2001 for a good review) and tends to conform to the requirements of the *in situ* approach to loess formation.

A major topic in the Russian literature is 'how did collapsibility develop?' This has never been discussed in 'western' literature because the reason for collapse is implicit in the aeolian deposition mechanism which forms loess deposits. In the Russian literature it appears to be the applied geomorphologists and ground engineers who continue to cling to the *in situ* theories of loess formation.

The problem of soil erosion falls within the purview of applied geomorphology. Because of its silty nature loess soil is very prone to erosion, by wind and water. The classic wind erosion events (e.g. the Dust Bowl) tend to be the blowing away of loess soil material; and the major water erosion problems are often loess connected – in particular the loss of soil material in north China. It is soil erosion that makes the Yellow River yellow. In north-west China where the loess is spread over mountainous terrain, there is a considerable landslide problem (Derbyshire *et al.* 2000). In 1920 a large earthquake caused much loss of life because of widespread collapse of loess tunnels which housed a large proportion of the population.

References

Derbyshire, E., Meng, X.M. and Dijkstra T.A. (eds) (2000) *Landslides in the Thick Loess Terrain of North-West China*, Chichester: Wiley.
Kriger, N.I. (1965) *Loess, Its Characteristics and Relation to the Geographical Environment*, Moscow: Nauka (in Russian).
——(1986) *Loess: Formation of Collapsible Properties*, Moscow: Nauka (in Russian).
Pecsi, M. (1990) Loess is not just the accumulation of dust, *Quaternary International* 7/8, 1–21.
Pye, K. (1987) *Aeolian Dust and Dust Deposits*, London: Academic Press.
Rozycki, S.Z. (1991) *Loess and Loess-like Deposits*, Wroclaw: Ossolineum.
Scheidig, A. (1934) *Der Loss und seine geotechnischen Eigenschaften*, Dresden and Leipzig: Steinkopf.
Smalley, I.J. (ed.) (1975) *Loess: Lithology and Genesis*, *Stroudsburg, PA*: Dowden, Hutchinson and Ross.
——(1980) *Loess – A Partial Bibliography*, Norwich: Geobooks (Elsevier).
Smalley, I.J. and Davin, J.E. (1980) The First Hundred Years: A Historical Bibliography of New Zealand loess 1878–1978, New Zealand Soil Bureau Bibliographical Report 28.
Trofimov, V.T. (ed.) (2001) *Loess Mantle of the Earth and its Properties*, Moscow: Moscow University Press (in Russian).

Further reading

Loess Letter ISSN 0110-7658; published twice a year by the International Union for Quaternary Research.
Collapsing Soils Communique ISSN 1473-0936; published by the Collapsing Soils Commission of the International Association of Engineering Geology and the Environment.
Loess Inform ISSN 0238-065X; published by the Hungarian Academy of Sciences; see in particular issue 2 for 1993 for general discussion.
Pecsi, M. (1968) Loess, in *Encyclopedia of Geomorphology*, 674–678, New York: Reinhold Book Corporation.
Pecsi, M. and Richter, G. (1996) Loss: Herkunft-Gliederung-Landschaften, *Zeitschrift für Geomorphologie, Supplementband* 98.

SEE ALSO: aeolian geomorphology; palaeoclimate; wind erosion of soil

IAN SMALLEY

LOG SPIRAL BEACH

There are several terms for the plan shape of asymmetrical beaches and bays, including log spiral, half-heart, crenulate, hook-shaped and zetaform. BEACHes between headlands consist of: a curved, nearly circular, portion in the lee of the headland, which may be absent in some areas; a logarithmic spiral section; and a nearly linear to curvilinear reach tangential to the downcoast headland. Yasso (1982) proposed that plan curvature is described by a logarithmic spiral law, with the distance from the centre of the spiral (r) to the beach increasing with the angle θ according to:

$$r = e^{\theta \cot \alpha}$$

where: θ is the angle of rotation, or spiral angle, which determines the tightness of the spiral; and

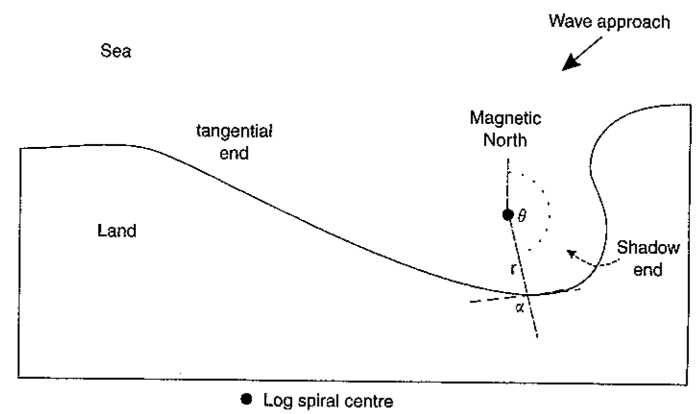

Figure 104 Log spiral beach and Yasso's (1982) logarithmic spiral law

α is the logarithmic spiral constant, the angle between a radius vector and the tangent to the curve at that point – this is constant for a given log spiral (Figure 104).

There may be systematic variations in beach sediment and morphology along log spiral beaches corresponding to longshore changes in wave height and energy. Beaches in California are finer grained and more gently sloping in the sheltered portions of log spiral bays, but they tend to be steep or reflective in southeastern Australia, while sections exposed to the dominant swell and storm waves are gentle or dissipative. The sheltered portion of mixed sand and coarse clastic bays in Alaska generally have low wave energy, small grain sizes, fairly well-sorted sediment, gentle beachface slopes and eroding shorefaces. The central portions have high wave energy, large grain sizes, poor sorting, moderate beachface slopes and shorelines transitional between erosion and deposition. The tangential ends of the bays are similar to the shadow ends, except that the beachfaces are steep and the shorelines depositional (Finkelstein 1982).

Progressive decline in beach curvature downdrift of headlands is usually thought to reflect increasing exposure to wave action, although wave energy is only one of the factors that must be considered. Shorelines try to attain equilibrium conditions determined by offshore wave refraction and diffraction, the distribution of wave energy, rates of longshore sediment transport and the relationships between beach slope, wave energy and grain size. Log spiral beaches formed by oblique waves are generally thought to be the most stable in nature. They are in static equilibrium when the tangential downcoast section is parallel to the wave crests, and as waves diffract into the bay they break simultaneously along the whole periphery. There is no longshore component of breaking wave energy and no littoral drift, and the plan shape, local beach slope and sediment size distribution are constant through time.

References

Finkelstein, K. (1982) Morphological variations and sediment transport in crenulate-bay beaches, Kodiak Island, Alaska, *Marine Geology* 47, 261–281.
Yasso, W.E. (1982) Headland bay beach, in M.L. Schwartz (ed.) *The Encyclopedia of Beaches and Coastal Environments*, 460–461, Stroudsburg, PA: Hutchinson Ross.

Further reading

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

ALAN TRENHAILE

LONG PROFILE, RIVER

A graph representing the relationship between altitude (H) and distance (L) along the course of a river, expressed by

$$H = f(L)$$

Any of three functions – exponential, logarithmic or power – can provide a reasonable fit to stream profiles and give rise to smooth, concave-upward curves (see GRADE, CONCEPT OF). Most long profiles tend to be concave, but they are not invariably smooth. Local steepening of channel gradient (see WATERFALL and KNICKPOINT) can result from such causes as more resistant bedrock strata, the introduction of a coarser or larger load, tectonic activity (see, for example, Riquelme *et al.* 2003) and BASE LEVEL changes (Knighton 1998) associated with REJUVENATION. Long profiles of glaciated valleys are often characterized by steps and overdeepenings and by hanging tributaries (MacGregor *et al.* 2000).

Some rivers have convex long profiles, and this may be a characteristic of updoming passive margins to continents or of discharge reductions downstream as is found in arid zone rivers (e.g. in the Namib).

Excessively or overconcave rivers occur in rivers with lower reaches that have become recently infilled by, for example, estuarine sedimentation or which have been affected by glacial diversion and subsequent lengthening of their courses (Wheeler 1979). At a local scale river long profiles may be punctuated by pool and riffle topography, and by the development of dune bedforms (see STEP-POOL SYSTEM).

The inverse relationship between channel gradient and discharge recognized by Gilbert (1877) helps to explain concavity, since tributary inflows cause a downstream increase of discharge which enables the stream's sediment load to be transported on progressively lower slopes. When discharge increases rapidly downstream with increasing contributing area, profile concavity is greater (Wheeler 1979). In addition, the calibre of sediment load is also related to stream gradient (see Richards 1982, for a review). However, whereas discharge is clearly an independent control of stream gradient, causation is less obvious in the gradient–sediment size relationship, for there are complex feedbacks between sediment size and gradient.

References

Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, Washington, DC: United States Geological and Geographical Survey.

Knighton, D. (1998) *Fluvial Forms and Processes, A New Perspective*, 2nd edition, London: Arnold.

MacGregor, K.R., Anderson, R.S., Anderson, S.P. and Waddington, E.D. (2000) Numerical simulations of glacial-valley longitudinal profile evolution, *Geology* 28, 1,031–1,034.

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

Riquelme, R., Martinod, J., Hérail, G., Darrozes, J. and Charrier, R. (2003) A geomorphological approach to determining the Neogene to recent tectonic deformation in the coastal cordillera of Northern Chile (Atacama), *Tectonophysics* 361, 55–275.

Wheeler, D.A. (1979) The overall shape of longitudinal profiles of streams, in A.F. Pitty (ed.) *Geographical Approaches to Fluvial Processes*, 241–260, Norwich: GeoAbstracts.

A.S. GOUDIE

LONGSHORE (LITTORAL) DRIFT

Longshore sediment transport is the net displacement of sediment parallel to a coastline. Such transport is maximized where WAVES (especially *breaking waves*) and wave-induced, shore-parallel, quasi-steady *longshore currents* (see CURRENT) prevail, i.e. in the *surf zone*. A secondary peak also occurs in the *swash zone*. Here wave energy is finally dissipated in the reversing currents of the wave *uprush* (flow at an angle to the shoreline under oblique wave approach) and *backwash* (flow normal to the shoreline under the force of gravity). This gives rise to a 'zig-zag' motion of water and sediment along the beach face. In both cases the obliqueness of the angle of wave approach is critical to the transport rates. The volume or mass of sediment transported by *longshore currents* and *swash* is termed *littoral drift*. Sediment may move as BEDLOAD, in continuous contact with the bed, or as SUSPENDED LOAD. In theory, particles in suspension are supported by turbulent (Reynolds) stresses within the fluid, while bedload particles are supported by inter-granular impact forces (Bagnold 1963). Debate exists concerning their relative importance. For example, Komar (1998) suggests that bedload comprises upwards of 75 per cent of the total transport, while Sternberg *et al.* (1989) suggest that in the surf zone, where waves are the pre-eminent entraining mechanism, suspended load can account for virtually all the longshore transport. The distinction between the two is, however, purely theoretical; the time-averaged mass concentration of particles decreases rapidly away from the immobile bed without a significant break, as the two support mechanisms merge. Further, since wave-induced, reversing oscillatory motions are the main entraining force, both suspended and bedload reveal strongly episodic behaviour. In the swash zone, the thin flows make the relative role of bedload and suspended load even more difficult to define. The variable directions of wave approach result in reversals of the direction of transport by both the longshore currents and the swash; thus, the *gross* sediment transport alongshore may be very large, while the *net* transport may be significantly smaller. However, transport rates up to several million $m^3 a^{-1}$ are common and this transport has been studied extensively by geomorphologists, coastal oceanographers and engineers concerned with coastal forms and dynamics, harbour siltation and dredging, the effectiveness of groynes as shore protection, etc.

Longshore sediment transport rates

Longshore transport is driven fundamentally by the alongshore component of the *wave momentum flux* or *radiation stress* (Longuet-Higgins and Stewart 1964), which itself creates a quasi-steady longshore current, although the complex interaction between the waves, the currents and the sediment is far from understood. Transport takes place within a combined *wave-current boundary layer*. The wave boundary layer is relatively thin compared to that of the current; consequently the stresses generated by the waves are significantly larger. However, to a first order, waves are purely oscillatory and thus cannot induce significant transport. It was this assumption, that waves entrain (initiate motion) and currents advect sediment, that led to the most common formulation for sediment transport alongshore (Bagnold 1963).

Longshore transport models

The Energetics Model: Inman and Bagnold (1963) defined longshore transport rates as a simple function of the alongshore component of the incident wave energy flux:

$$I_l = (\rho_s - \rho_f)\, ga' Q_l = K\, P_L$$
$$= K\,(E\, C\, n)_b \sin \alpha_b \cos \alpha_b$$

where I_l is immersed weight transport rate, a′ is a constant of 0.6, Q_l is longshore transport rate, ρ_s and ρ_f are solid and fluid mass densities, K is a dimensionless proportionality coefficient, E is the specific wave energy density, C is the wave celerity or speed, n is the ratio of wave group velocity to wave phase velocity, α is the angle of wave breaking and the subscript b refers to conditions at the point of wave breaking. The constant K, which was originally proposed as an efficiency factor (≈ 0.77 when the root-mean-square wave height is used in the energy calculation; Komar 1971), has subsequently been the source of considerable debate. This proportionality factor is also thought to be dependent upon the grain size, wave steepness, the surf similarity parameter or Irribarren Number, bed slope, etc. The 'constant' K was originally conceived as applying to fully developed transport in an instantaneous sense. A number of authors have used a time-averaged relationship to predict the potential for a time-averaged rate of transport and thus a model for the long-term potential for erosion and deposition (see Greenwood and McGillivray 1980). Komar (1998) gives an extensive review of the energetics model.

Bailard (1981), following Bagnold (1966) and Bowen (1980), expanded the basic model and showed that the total longshore transport rate depends upon the steady current and a number of higher order moments of the velocity field:

$$<i_y> = \rho\, c_f u_m^3 \left\{ \left[\frac{\varepsilon_b}{\tan \phi}\right] \left(\delta_v^3 + \frac{\delta_v}{2} + \frac{\tan \beta}{\tan \phi} (u3)^* \tan \alpha \right) + \frac{u_m}{w} \varepsilon_s \delta_v (u3)^* + \frac{u_m^2}{w^2} \varepsilon_s^2 \tan \beta\, (u5)^* \tan \alpha \right\}$$

where c_f is a drag coefficient, u_m is oscillatory velocity magnitude, ϕ is angle of internal friction, α is angle of wave approach, ε_b is bedload efficiency, ε_s is suspended load efficiency, w is sediment fall velocity and the relatively steady currents δ, δ_u and δ_v are defined as:

$$\delta = \frac{\bar{u}}{u_m}\ ;\ \delta_u = \frac{\bar{u}}{u_m} \cos \theta\ ;\ \delta_v = \frac{\bar{u}}{u_m} \sin \theta$$

and the velocity moments ψ_1 and ψ_2 are defined as:

$$\psi_1 = \frac{<\tilde{u}^3>}{u_m^3}\ ;\ \psi_2 = \frac{<|\overset{\rightharpoonup}{u}|^3 \tilde{u}>}{u_m^4}$$

and the integrals $(u3)^*$ and $(u5)^*$ contain the time-averaged magnitudes of the combined velocities associated with each of the higher moments and are defined as:

$$(u3)^* = \frac{1}{T}\int_0^T (\delta^2 + 2\delta\cos(\theta-\alpha)\cos\sigma t + \cos^2\sigma t)^{\frac{3}{2}}\,dt$$

$$(u5)^* = \frac{1}{T}\int_0^T (\delta^2 + 2\delta\cos(\theta-\alpha)\cos\sigma t + \cos^2\sigma t)^{\frac{5}{2}}\,dt$$

where T is the wave period, θ is steady current angle and σ is wave frequency.

The Applied Stress Model uses the concept of 'excess stress' for alongshore transport, i.e. the stress in excess of that used to initiate motion of the sediment. The models use precise physical relationships coupled with semi-empirical expressions to describe the transport process. A typical model of the longshore component of transport, q_l, is that of Grant and Madsen (1979):

$$q_l(t) = 40wD\left(\frac{\frac{1}{2}\rho c_f\left(u^2(t)+v^2(t)\right)}{(s-1)\rho g D}\right)^3 \frac{v(t)}{\sqrt{u^2(t)+v^2(t)}}$$

where t is time, w is mean sediment fall velocity, D is sediment diameter, ρ is fluid density, c_f is coefficient of bed friction, u and v are horizontal cross-shore and longshore fluid velocities, s is specific gravity and g is gravitational constant.

Most of the models have been calibrated using sand-sized material, but clearly longshore transport also includes coarser grained materials. Van Wellen *et al.* (2000) reviews the longshore transport equations for coarse-grained beaches.

Geomorphological significance

Littoral drift is an important part of the *sediment budget*, which is based on mass conservation principles applied to the coastal zone. A major concept in coastal geomorphology is that of the littoral cell, consisting of a zone of sediment supply (may be rivers or shore erosion), a zone of erosion (net loss of sediment), a zone of longshore transport and a zone of accretion (net gain of sediment). An excellent review of the concept is available in Carter (1988). *Gradients* in the rates of longshore sediment transport can be used to model major aspects of shoreline erosion and/or accretion (see Komar 1998).

Several specific geomorphological forms owe their origin to longshore sediment transport. For example, a TOMBOLO is a strip of sediment accumulating between an offshore island and the main shoreline formed as a result of wave refraction around the island. Refraction produces waves, which approach from opposing directions in the lee of the island and induce currents and longshore transport, which converge from two directions. Spectacular *sand spits* may develop as a result of littoral drift being deposited at a re-entrant in the shoreline or across embayments and may stretch for several kilometres. Refraction of waves around the down drift terminus often produce *recurved* or *hooked* spits.

References

Bagnold, R.A. (1963) Mechanics of Marine Sedimentation, in M.L. Hill (ed.) *The Sea: Ideas and Observations*, Vol. 3, 507–528, New York: Wiley Interscience.

——(1966) An approach to the sediment transport problem from general physics, *United States Geological Survey Professional Paper* 422-I.

Bailard, J.A. (1981) An energetics model for a plane sloping beach, *Journal of Geophysical Research* 86, 10,938–10,954.

Bowen, A.J. (1969) The generation of longshore currents on a plane beach, *Journal of Marine Research* 27, 206–214.

——(1980) Simple models of nearshore sedimentation: beach profiles and longshore bars, in S.B. McCann (ed.) *The Coastline of Canada*, 1–11, Ottawa: Geological Survey of Canada.

Carter, R.W.G. (1988) *Coastal Environments. An Introduction to the Physical, Ecological and Cultural Systems of Coastlines*, London: Academic Press.

Grant, W.D. and Madsen, O.S. (1979) Combined wave current interaction with a rough bottom, *Journal of Geophysical Research* 84, 1,797–1,808.

Greenwood, B. and McGillivray, D.G. (1980) Modelling the impact of large structures upon littoral transport in the Central Toronto waterfront, Lake Ontario, Canada, *Zeitschrift für Geomorphologie Supplementband* 34, 97–110.

Inman, D.L. and Bagnold, R.A. (1963) Littoral processes, in M.L. Hill (ed.) *The Sea: Ideas and Observations*, Vol. 3, 529–543, New York: Wiley Interscience.

Komar, P.D. (1971) The mechanics of sand transport on beaches, *Journal of Geophysical Research* 76, 713–721.

——(1998) *Beach Processes and Sedimentation*, Upper Saddle River, NJ: Prentice Hall.

Longuet-Higgins, M.S. and Stewart, R.W. (1964) Radiation stresses in water waves: a physical discussion with applications, *Deep Sea Research* 11, 529–562.

Sternberg, R.W., Shi, N.C. and Downing, J.P. (1989) Suspended sediment measurements, in R.J. Seymour (ed.) *Nearshore Sediment Transport*, 231–257, New York: Plenum.

Van Wellen, E., Chadwick, A.J. and Mason, T. (2000) A review and assessment of longshore transport equations for coarse-grained beaches, *Coastal Engineering* 40, 243–275.

Further reading

Bayram, A., Larson, M., Miller, H.C. and Kraus, N.C. (2001) Cross-shore distribution of longshore sediment transport: comparison between predictive formulas and field measurements, *Coastal Engineering* 44, 79–99.

Komar, P.D. and Inman, D.L. (1970) Longshore sand transport on beaches, *Journal of Geophysical Research* 75, 5,914–5,927.

Longuet-Higgins, M.S. (1970a) Longshore currents generated by obliquely incident waves, 1, *Journal of Geophysical Research* 75, 6,778–6,789.

——(1970b) Longshore currents generated by obliquely incident waves, 2, *Journal of Geophysical Research* 75, 6,790–6,801.

——(1972) Recent progress in the study of longshore currents, in R.E. Meyer (ed.) *Waves on Beaches and Resulting Sediment Transport*, New York: Academic Press.

Van Rijn, L.C. (1998) *Principles of Coastal Morphology*, Amsterdam: Aqua Publications.

BRIAN GREENWOOD

LUNETTE

Transverse and roughly concentric aeolian accumulations (the *bourrelets* of some French workers) that occur on the downwind margins of PANS. Although they had been described before, they were named as such in Australia by Hills (1940), though the basis of his etymology is unclear. They tend to occur in areas where present-day precipitation levels are between about 100 and 700 mm, but their stratigraphy can give a good indication of past changes in climate and hydrological conditions. Some basins may have two or more lunettes on their lee sides and these may have different grain size and mineralogical characteristics. Lunettes may be some kilometres long and in exceptional circumstances may attain heights in excess of 60 m.

Good regional descriptions of lunettes are provided for the High Plains of the USA by Holliday (1997), for Tunisia by Perthuisot and Jauzein (1975), for the Kalahari by Lancaster (1978) and for the Pampas of Argentina by Dangavs (1979).

The materials that make up lunettes can vary from clay-sized material (which in the case of clay dunes (Bowler 1973) can make up 30 to 70 per cent of the total) through to sand-sized material. Equally some lunettes are carbonate-rich (Goudie and Thomas 1986) whereas others are almost pure quartz. Lunettes may also contain appreciable quantities of evaporite minerals derived from the basins to their windward.

Various hypotheses have been put forward to explain lunette composition. Hills (1939) believed that the lunettes were built up when the pans contained water and that they were composed of atmospheric dust captured by spray droplets derived from the water body. Stephens and Cocker (1946) pointed out that this could not account for those lunettes that were not predominantly silty. They also suggested that many of the lunettes were built up of aggregates transported from the floors of pans. Campbell (1968) believed that this deflation hypothesis could indeed account for many lunette features. As she remarked (p. 104) 'the close similarity between the composition of the lunette and its associated lake bed suggested that the two are causally related, i.e. that the material in the lunette was derived from the lake bed'. However, she also recognized that some of the material could be derived from wave-generated beaches and so could be analogous to primary coastal foredunes.

This was a view that was developed by Bowler (1973), who saw sandy facies as being associated with a beach provenance (at times of relatively high water levels) whereas clay-rich facies, which also may have a high content of evaporite grains, formed during drier phases when deflation of the desiccated lake floor was possible. Lunettes can therefore provide evidence for understanding past hydrological changes (Page *et al.* 1994).

References

Bowler, R.M. (1973) Clay dunes: their occurrence, formulation and environmental significance, *Earth-Science Reviews* 9, 315–338.

Campbell, E.M. (1968) Lunettes in South Australia, *Transactions of the Royal Society of South Australia* 92, 83–109.

Dangavs, N.V. (1979) Presence de dunes de argilla fosiles en La Pampa Deprimada, *Revista Association Geologica Argentina* 34, 31–35.

Goudie, A.S. and Thomas, D.S.G. (1986) Lunette dunes in Southern Africa, *Journal of Arid Environments* 10, 1–12.

Hills, E.S. (1939) The physiography of north-western Victoria, *Proceedings of the Royal Society of Victoria* 51, 297–323.

Hills, E.S. (1940) The lunette: a new landform of aeolian origin, *Australian Geographer* 3, 1–7.
Holliday, V.T. (1997) Origin and evolution of lunettes on the High Plains of Texas and New Mexico, *Quaternary Research* 47, 54–69.
Lancaster, I.N. (1978) The pans of the Southern Kalahari, Botswana, *Geographical Journal* 144, 80–98.
Page, K., Dare-Edwards, A., Nanson, G. and Price, D. (1994) Late Quaternary evolution of Lake Urana, New South Wales, Australia, *Journal of Quaternary Science* 9, 47–57.
Perthuisot, J.P. and Jauzein, A. (1975) Sebkhas et dune d'argile: l'enclave endoréique de Pont du Fahs, Tunisie, *Revue de Géographie Physique et de Géologie Dynamique* 17, 295–306.
Stephens, C.G. and Crocker, R.L. (1946) Composition and genesis of lunettes, *Transactions of the Royal Society of South Australia* 70, 302–312.

A.S. GOUDIE

M

MAGNITUDE–FREQUENCY CONCEPT

When the activity of different geomorphic processes is compared on a given timescale some processes appear to operate continuously while others operate only when specific conditions occur (referred to as events). The term eposidicity refers to the tendency of processes to exhibit discontinuous behaviour and occur sporadically as a series of individual events. Episodicity occurs when discontinuity is inherent in the forcing process (e.g. discontinuous rainfall produces episodic gully erosion). It may also occur if the relationship between the forcing process and the geomorphic response is not constant (e.g. the strain resulting from continuous crustal plate convergence may be manifest as either continuous deformation or episodic earthquake activity and coseismic uplift). The point at which the resistance to stress imposed by a forcing process is overcome is marked by a discernible geomorphic response (event) and is referred to as a threshold (see THRESHOLD, GEOMORPHIC). Because of hysteresis effects, the initiating threshold for a geomorphic event may be of a different magnitude to the terminating threshold.

Like many concepts in geomorphology, interpretations of eposidicity are scale dependent. Even processes that are sometimes considered continuous can be interpreted as episodic on different timescales. For example, SOIL CREEP in a given area is commonly portrayed as continuous and ubiquitous in terms of landform evolution. However, on a diurnal or seasonal scale, some forms of soil creep are evidently episodic, operating only when certain temperature or moisture conditions are met. For a given process, large events involving high amounts of concentrated energy (high magnitude) are rare (low frequency) and small events are common and tend to be less episodic.

Historically geomorphology has gone through a number of major paradigm shifts involving the eposidicity and the effectiveness of processes. The earliest interpretations of how landforms were created were based on catastrophic formative supernatural or natural events unrelated to the work of contemporary processes. As the science of geomorphology developed, increasing recognition of the age of the Earth allowed for the possibility that slow acting, contemporary processes, with sufficient time, could account for the development of much of the form evident in the landscape. Paradoxically the ability to date accurately key deposits revealed that certain landforms had been formed under specific and unusual circumstances. This gave rise to the prospect of NEOCATASTROPHISM. The consequential question that heightened interest in the study of magnitude–frequency relations was: were the effects of a rare catastrophic event 'permanent' or, given time, were they overwhelmed by 'normal', small but constantly operating processes responsible for producing a characteristic form? The prospect of identifying a 'characteristic' form in a geomorphic system however requires stability in boundary conditions such as climate, tectonic activity and vegetation cover and thus the existence of some form of DYNAMIC EQUILIBRIUM. Clearly the shorter the period of observation the more likely it is that these conditions pertain.

From the point of view of landform development, the RELAXATION TIME associated with a geomorphic event may be a more appropriate parameter than magnitude for identifying the formative or geomorphically effective event. The relaxation time is the length of time over which the effects of an event can be discerned in the landscape. Although this is clearly a function of event magnitude it also incorporates the influence of terrain resilience, LANDSCAPE SENSITIVITY and ambient conditions (Crozier 1999). The degree of equilibrium and the prospect of identifying characteristic form can then be determined by the 'transient form ratio' (Brundsen and Thornes 1979) which is the ratio of relaxation time of a geomorphic event to its recurrence interval (frequency). Values equal or greater than one indicate a constantly changing system while values less than one represent some form of dynamic equilibrium.

Many studies on this question, including the seminal work by Wolman and Miller (1960) have drawn conclusions based only on the period of instrumental record and the extrapolation of such results to landform evolution must be done with care (Wolman and Gerson 1978). In Wolman and Miller's original study it was shown that most of the work of sediment transportation in rivers was carried out by moderate flow events of a magnitude that recurs on average at least once every five years. The analytical approach leading to this conclusion is illustrated in Figure 105. Curve (a) indicates that the rate of sediment movement is a power function of applied stress, or in this case a surrogate such as discharge. Curve (b) shows a log

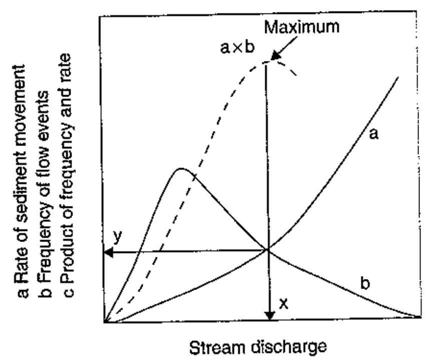

Figure 105 Relations between rate of sediment transport, discharge and the frequency–magnitude distribution of discharge events, c= a × b (after Wolman and Miller 1960)

normal frequency–magnitude distribution of measured flow events. Curve (a × b) is the product of frequency and rate of movement which attains a maximum work value which can be related back to the discharge axis at (x) to identify the magnitude of the event performing the most work. The frequency of this event (y) can be determined with reference to the frequency–magnitude distribution curve. Wolman and Miller's work has been extended to a number of different processes and environments, in some cases producing conclusions at variance with the original findings. Selby (1974), for example, argues that in high-energy hillslope environments and the headwaters of drainage basins, high magnitude landslide events occurring at frequencies less than once every five years dominate geomorphic activity on slopes and low-order stream channels. Similar conclusions have been made by a comprehensive comparison of fluvial transport and mass movement in different parts of the drainage basins (Trustrum *et al.* 1999).

In strictly geomorphic terms, the magnitude of an event generally refers to the amount of work carried out (e.g. mass of sediment transported) or the degree of landform change experienced (geomorphic response). However, because of difficulty in directly measuring geomorphic events many magnitude–frequency studies are approached indirectly. This involves characterizing events from the behaviour of the forcing agent, rather than from the geomorphic event itself. However, the indirect approach implies a known relationship between geomorphic work and the behaviour of the forcing agent. For example, if this relationship and the initiating thresholds are known, the frequency and magnitude of soil erosion and landsliding can be determined from the rainfall record, aeolian transport from wind speed, coastal changes from the wave regime and fluvial transport from the streamflow regime.

A major problem with the indirect approach is that relationships between the forcing process and geomorphic work are not always close or temporally or spatially stable. For example, stream discharge may not relate closely to sediment transport in a supply-constrained system. Indeed in some streams, sediment load may be related to stream power at certain times and to rates of hillslope sediment supply at others. The nature of sediment control can also change throughout the catchment to the extent that sediment load in steep upper catchments may relate

to the magnitude and frequency of landslide events. Furthermore, the parameter used to represent magnitude also needs to be carefully chosen if it is to reflect accurately and consistently the energy of the forcing process. In the case of landsliding for example, the groundwater content of a hillslope is directly responsible for initiating movement, however lack of adequate databases dictates that magnitude–frequency analysis is often carried out using rainfall values; a more appropriate analysis would involve the magnitude and frequency of groundwater levels. The use of arbitrarily defined events such as daily rainfall as opposed to event rainfall introduces another level of inaccuracy into the analysis, as important parameters such as actual intensity and duration are obscured. Besides magnitude and frequency, duration is another parameter of the forcing processes that influences geomorphic response. Clearly, in power-constrained systems, high effective streamflow, maintained for long periods of time, will accomplish more work than the same levels of flow occurring as short duration events. Similarly, studies of the relationship between the magnitude and intensity of earthquakes and the degree of land deformation have shown that the duration of shaking is important in evoking a geomorphic response.

For a given process, average event frequencies can be established from the number of times an event initiating threshold is exceeded or the number of times the threshold is exceeded by a specified magnitude, in a unit of time. More commonly, following the practice of flood frequency analysis, frequencies are expressed as recurrence intervals (return periods) determined by the relationship of the length of record to the ranking or relative magnitude of the event in a series of events. A range of statistical models for extreme value probability distributions can be used for depicting the declining exponential function between magnitude and frequency. Little confidence can be placed on derived event magnitudes with recurrence intervals greater than the length of record.

An important property of recurrence intervals (sometimes overlooked when regional comparisons are made) is that they are generally a function of the size of the area from which they are derived. In other words, the larger the area, all other things being equal, the more likely it will be to experience a specified event and therefore the shorter the derived recurrence interval. Another point of caution is that the statistical derivation of frequency may obscure the clustering of events, which is a behavioural property that can be significant in generating a geomorphic response. Because variation of frequency through time may signal significant environmental change the identification of clustering is an important aspect of magnitude–frequency analysis.

For some processes, magnitude and frequency of occurrence in space (Innes 1985; Hovius *et al.* 1997) mirrors temporal magnitude–frequency relationships. In the case of landslides for example, episodes producing numerous landslides on the one occasion show that large landslides are greatly outnumbered by smaller landslides. These spatial magnitude–frequency relationships are used in hazard assessments as analogues for temporal magnitude–frequency relationships. Establishing magnitude–frequency relationships from landscape evidence needs to take account of the longevity of landscape evidence; clearly the signature of large landslides persists longer than that of smaller landslides.

The magnitude–frequency concept and the analysis it has provoked have been important in informing the discipline about the variability and behaviour of geomorphic processes. It reminds us that within one lifetime we are unlikely to experience the whole range in magnitudes that a particular process is capable of generating. The concept provides a rationale for extrapolating short-term measurements over longer periods, as a way of assessing the long-term rates of geomorphic processes. However changes in boundary conditions limit the extent to which the relationships can be extrapolated. Magnitude–frequency analysis also provides a method for statistically identifying the key events in terms of work and landform response, thereby providing a key variable for characterizing geomorphic systems and predicting other system characteristics. Finally, from a pragmatic point of view, the concept enables the identification of a design or planning event for use in engineering decisions and hazard management.

References

Brunsden, D. and Thornes, J.B. (1979) Landscape sensitivity and change, *Transactions Institute British Geographers* 4, 463–484.

Crozier, M.J. (1999) The frequency and magnitude of geomorphic processes and landform behaviour, *Zeitschrift für Geomorphologie N.F. Supplementband* 115, 35–50.

Hovius, N., Stark, C.P. and Allen, P.A. (1997) Sediment flux from a mountain belt derived from landslide mapping, *Geology* 25, 231–234.
Innes, J.L. (1985) Magnitude-frequency relations of debris flows in northwest Europe, *Geografiska Annaler* 67A, 23–32.
Selby, M.J. (1974) Dominant geomorphic events and landform evolution, *Bulletin International Association Engineering Geologists* 9, 85–89.
Trustrum, N.A., Gomez, B., Page, M.J., Reid, L.M. and Hicks, D.M. (1999) Sediment production and output: the relative role of large magnitude events in steepland catchments, *Zeitschrift für Geomorphologie N.F. Supplementband* 115, 71–86.
Wolman, M.G. and Gerson, R. (1978) Relative scales of time and the effectiveness of climate in watershed geomorphology, *Earth Surface Processes and Landforms* 3, 189–208.
Wolman, M.G. and Miller, J.P. (1960) Magnitude and frequency of forces in geomorphic processes, *Journal of Geology* 68, 54–74.

Further reading

Crozier, M.J. and Mausbacher, R. (eds) (1999) Magnitude and frequency in geomorphology, *Zeitschrift für Geomorphologie N.F. Supplementband* 115.
Selby, M.J. (1993) *Hillslope Materials and Processes*, 2nd edition, Oxford: Oxford University Press.

MICHAEL J. CROZIER

MANAGED RETREAT

Usually defined as the backward realignment of coastal defences so as to allow the reformation of SALTMARSH or mud flat (see MUD FLAT AND MUDDY COAST) on the seaward side. It may also, rarely, be used where other coastal features such as dunes are expected to form in front of a retreating line of defence. The term has now largely been superseded by the preferred phrase '*managed realignment*' which is effectively synonymous but avoids the negative implications to many people of the word 'retreat'.

In low-lying tidally dominated coastal areas, especially in northern Europe and North America, large areas of saltmarsh have been 'reclaimed' by the building of sea walls for agricultural use or for urban and industrial development. Such land claim alters the dynamics of the local tidal flow and sediment budget and may lead to localized increase in tidal energy and to EROSION. Reclaimed areas are often in estuaries; the natural funnel shape of an ESTUARY is transformed to a narrower more parallel one providing less space for the dissipation of tidal energy. The tidal flow tends to move further upriver, the river channel deepens and widens and the mudflat/saltmarsh complex migrates also, maintaining its relation to the local energy gradient (Pethick 2000). Although seawalls were often originally positioned so as to leave in front some marsh and mudflat, these often become eroded away. High tides come up to the sea defences and the threat of flooding at spring tides and during storms increases; the walls themselves may be destroyed. If also SEA LEVELS are rising, the effect is magnified.

Appropriate responses to the threat depend on the type of infrastructure in the hinterland. Raising and strengthening seawalls, while expensive, may be thought the best option if valuable built-up or agricultural land is protected. Where the hinterland consists of low value agricultural land, managed retreat is seen as the best solution to the erosion problem. Reinstatement of the coastal marshes should provide a dispersal area for the tidal flow, a supply of sediment for local coastal processes, and a buffer zone for dispersion of wave and tide energy. Given time and space, it is hoped that the coastline will develop its own contours and maintain itself, altering to suit changing conditions without the need for expensive engineering works.

However, breaching or removal of seawalls does not always lead to the reestablishment of saltmarsh. Some low-level sites inundated either through storm damage or by the deliberate realignment of walls have remained as largely unvegetated mudflats (French *et al.* 2000) over a period of as much as fifty years. While popular with environmentalists and relatively inexpensive compared with hard engineering solutions, managed realignment is less attractive to the owners of the land to be sacrificed.

The success of realignment programmes has been shown to depend on several factors. It appears only to be successful on sites where saltmarsh once existed and where marshes survive nearby as sources of suitable plant propagules. There should be traces of the original saltmarsh creek system (though a channel system may initially be artificially created). The site should be sufficiently high that it is only inundated 400 to 450 times annually – in practice this means above 2.1 m OD or only mud flat is likely to form. It should preferably be neither completely flat nor steeply sloping but with a gentle land to sea slope (Burd 1995).

At least in the EU, managed realignment schemes can attract funding under the saltmarsh option of the Countryside Stewardship scheme (which replaced the Habitat Scheme in January 2000). Detailed Coastal Habitat Management Plans need to be drawn up in all cases before action is taken.

References

Burd, F. (1995) *Managed Retreat: A Practical Guide*, Peterborough: English Nature.
French, C.E., French, J.R., Clifford, N.J. and Watson, C.J. (2000) Sedimentation-erosion dynamics of abandoned reclamation; the role of waves and tide, *Continental Shelf Research* 20 (12–13), 1,711–1,733.
Pethick, J. (2000) Coastal management and sea level rise, *Catena* 42, 307–322.

Further reading

French, P.W. (2001) *Coastal Defences: Processes, Problems and Solutions*, London: Routledge.

PIA WINDLAND

MANGROVE SWAMP

Mangroves are trees or shrubs which grow in sheltered, or low-energy, upper intertidal zones in the tropics and subtropics, replacing SALTMARSH which is found along muddy temperate shorelines. There are more than fifty species of mangroves occurring in two distinct provinces, the Indo-west Pacific which contains the greatest diversity of species, and the West Indian region with far fewer species. In the West Indies, three species of mangrove occur, each tending to occupy a discrete location, *Rhizophora mangle* to seaward, *Avicennia germinans* in more landward locations, and *Laguncularia racemosa* mixed with the other species, or in areas that have been disturbed. In the Indo-west Pacific province there are more species of mangroves, with up to thirty species in the most diverse locations. There is generally a seaward zone, composed of species of *Avicennia* or *Sonneratia*, an intermediate zone dominated by species of *Rhizophora* and *Bruguiera*, and more landward zones in which *Ceriops* occurs with genera such as *Lumnitzera, Heritiera* and stunted forms of *Avicennia*. In areas of heavy rainfall mangrove forests merge into tropical forests, whereas in arid regions they often contain a hypersaline mudflat and are flanked by bare or samphire-covered supratidal flats.

Mangroves can grow in freshwater, but they appear to have a competitive advantage over other vegetation when the substrate is saline. Mangroves perform best at salinities that are less than that of seawater, but can survive in salinities of up to 90 parts per thousand. They show a series of adaptations to saline sediments. The root systems of mangroves contain breathing cells called lenticels, and are often distinctive, enabling the plants to survive in upper intertidal habitats which are subject to frequent inundation and are composed of muds or sandy muds which are anaerobic. The enormous prop root systems that characterize *Rhizophora* (Plate 75) are almost impenetrable. *Avicennia* and *Sonneratia* have pencil-like roots termed pneumatophores, whereas other species have knee or buttress roots. In addition many species of mangroves produce viviparous seedlings, the fruit already producing a well-developed root before falling from the tree.

These adaptations to the inhospitable environment in which mangroves live, and the zonation of species which is apparent on many shorelines, led to an early interpretation that mangroves undergo a succession of species culminating in a terrestrial climax vegetation. The root systems slow the movement of water and may trap sediment promoting accretion along the shore. There have been various attempts to measure sedimentation rates beneath mangrove forests. Direct measurements have been made by using systems

Plate 75 The interior of a mangrove swamp in the West Indies. The mangrove *Rhizophora mangle* has a network of prop roots that extend two or more metres up the trunks of the trees. These root systems can substantially reduce the flow of water across the substrate encouraging deposition of mud where suspended in the water column. In other cases, such as the forest shown here, the fibrous roots themselves contribute to a mangrove peat substrate

of stakes throughout the mangroves, or placing a marker layer. These approaches have generally met with less success than in saltmarshes because of the active bioturbation of the muds by a rich and diverse fauna, including crabs and mud skippers. Bioturbation also limits the efficacy of determining sedimentation rates using ^{210}Pb or ^{137}Cs isotopes, but rates of up to $3\,mm\,yr^{-1}$ are indicated (Lynch *et al.* 1989). Longer term sedimentation rates have been determined using radiocarbon dating and indicate that rates of up to $6–8\,mm\,yr^{-1}$ can occur. However, the short-term and long-term estimates may not be directly comparable because it is often difficult to discriminate whether root material was formed close to, or at some depth below, the surface. It is also difficult to take compaction into account, or to allow for the variation in sedimentation rate that is likely across the intertidal zone.

Zonation of mangrove species, even if it does occur, need not indicate a temporal succession that involves replacement of successive zones through time. Patterning of species may be a static equilibrium in relation to environmental gradients in habitat factors, such as salinity or water-logging (Lugo 1980). It has proved useful to distinguish various geomorphologically defined habitats within which mangroves grow (Thom *et al.* 1975; Semeniuk, 1985).

There is considerable variation both in the mangrove species that are found, and the elevation at which they grow, up estuaries. Particularly extensive mangrove forests occur in the abandoned parts of deltas where tidal processes are important, such as the Sundarbans west of the mouth of the Ganges–Brahmaputra rivers. Where mangrove swamps are associated with the mouths of large rivers, the substrate on which they are established is generally composed of terrigenous sediments washed from the catchments. The patterning of mangrove species in such deltaic environments reflects an ever-changing series of geomorphological habitats in which mangroves respond opportunistically to habitat change induced by geomorphological processes, such as channel migration and avulsion (Thom 1967).

Mangroves also occur in reef environments where their extent is often a function of stage of reef evolution (Stoddart 1980). In these settings, mangroves either establish over calcareous sediments formed from reef organisms, or they develop over mangrove peat which is derived from the roots of the mangroves themselves. Mangrove islands within the Belize barrier reef, termed mangrove ranges (Plate 76), are complex. Although these islands are dominated by tidal overwash, the patterning of species is not straightforward, with *Rhizophora* adopting a dwarfed scrub form in some locations, and elsewhere reaching a woodland of several metres tall, intergrading into areas of more open *Avicennia* woodland. In places the mangrove forests are dissected by sinuous creeks, and elsewhere there are bare areas which may result from storm damage, but where soil chemistry presently appears to prevent recolonization by mangroves.

It is clear that mangrove swamps have altered in extent considerably. The stratigraphy of mangrove-dominated coasts records the Holocene evolution of coastal environments and there have been significant adjustments as a result of changes in sea level. Throughout much of the Indo-west Pacific region, where sea level achieved a level close to present around 6,000 years ago and has been relatively stable or fallen slightly since, former mangrove sediments often underlie extensive Holocene plains upon which freshwater wetlands or peat swamp forest have established (Woodroffe *et al.* 1985). Former mangrove sediments often represent potential acid-sulphate soils in which pyrite can become oxidized, resulting in extremely acidic waters if drained or excavated. In the West Indies and the Everglades of Florida, where the history of relative SEA LEVEL change through the Holocene appears to have been characterized by rise to present, intertidal mangrove peat overlies previously terrestrial environments, such as freshwater sedge peat.

Plate 76 A mangrove island, called a mangrove range, on the Belize barrier reef. Although dominated by two mangrove species, *Rhizophora mangle* and *Avicennia germinans*, there is a complex vegetation pattern with areas that are bare of vegetation and a network of creeks that dissect the island

These productive and protective forests are also likely to undergo changes in distribution as a result of sea-level change in the future (Woodroffe 1990). Although this has attracted considerable attention, it is the case that human impact, clearing forests for other land use, or for timber, and most recently for shrimp farming, have generally already had far-reaching impacts. Mangroves are subject to various natural disturbances, for instance storm impact. However, they play an important role as storm protection, especially where surges are experienced as in the Bay of Bengal.

References

Lugo, A.E. (1980) Mangrove ecosystems: successional or steady state? *Biotropica (Supplement)* 12, 65–95.

Lynch, J.C., Meriwether, J.R., McKee, B.A., Vera-Herrera, F. and Twilley, R.R. (1989) Recent accretion in mangrove ecosystems based on ^{137}Cs and ^{210}Pb, *Estuaries* 12, 284–299.

Semeniuk, V. (1985) Development of mangrove habitats along ria shorelines in north and northwestern tropical Australia, *Vegetatio* 60, 3–23.

Stoddart, D.R. (1980) Mangroves as successional stages, inner reefs of the northern Great Barrier Reef, *Journal of Biogeography* 7, 269–284.

Thom, B.G. (1967) Mangrove ecology and deltaic geomorphology: Tabasco, Mexico, *Journal of Ecology* 55, 301–344.

Thom, B.G., Wright, L.D. and Coleman, J.M. (1975) Mangrove ecology and deltaic-estuarine geomorphology, Cambridge Gulf-Ord River, Western Australia, *Journal of Ecology* 63, 203–222.

Woodroffe, C.D. (1990) The impact of sea-level rise on mangrove shorelines, *Progress in Physical Geography* 14, 483–520.

Woodroffe, C.D., Thom, B.G. and Chappell, J. (1985) Development of widespread mangrove swamps in mid-Holocene times in northern Australia, *Nature* 317, 711–713.

Further reading

Alongi, D. and Robertson, A. (eds) (1992) *Tropical Mangrove Ecosystems*, Washington, DC: American Geophysical Union, Coastal and Estuarine Studies.

Hogarth, P.J. (1999) *The Biology of Mangroves*, Oxford: Oxford University Press.

COLIN WOODROFFE

MANTLE PLUME

'A blob of relatively hot, low-density mantle that rises because of its buoyancy' (Condie 2001: 1). Their existence was suggested by J.T. Wilson (1963) who sought to explain the progressive change in age and form of islands along oceanic chains, such as the Hawaiian–Emperor Chain of the Pacific. He proposed that as a lithospheric plate moves across a fixed hotspot (the mantle plume), volcanic activity is recorded as a linear array of volcanic seamounts and islands parallel to the direction in which the plate moves (Plate 77).

As rising mantle plumes reach the base of the lithosphere, they spread laterally to produce plume heads which may have diameters that reach 500 to 3,000 km. The tails of the plumes are only 100–200 km in diameter. The surface manifestations of plumes are large hotspots, which are zones with active volcanism. Those plumes that are 1,500–3,000 km in diameter are termed superplumes (Condie 2001: 2).

Mantle plumes are of fundamental geomorphological importance. First, their associated hotspots help to explain the distribution of volcanic activity in intraplate situations. They also explain the development of volcanic chains on the Pacific plate and elsewhere, and also the development of some seamounts. Second, broad zones of uplift (swells) are sometimes associated with mantle plumes. These can create large domes that then dominate drainage patterns at a regional and subcontinental scale. Cox (1989) argued that the drainage directions and patterns of India, eastern South America and southern Africa were good examples of this effect (Figure 106). Third, many large igneous provinces (LIPs) have a mantle

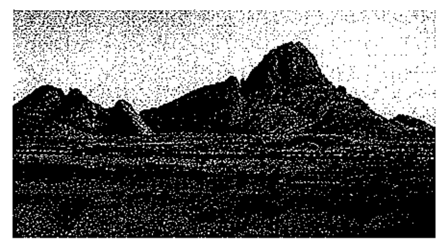

Plate 77 The Spitzkoppje in central Namibia is a granite mass that was emplaced in the early Cretaceous as a result of magmatic activity associated with the presence of a mantle plume that contributed to the rifting of Gondwanaland. The hotspot is now located in the Southern Ocean in the vicinity of Gough and Tristan da Cunha

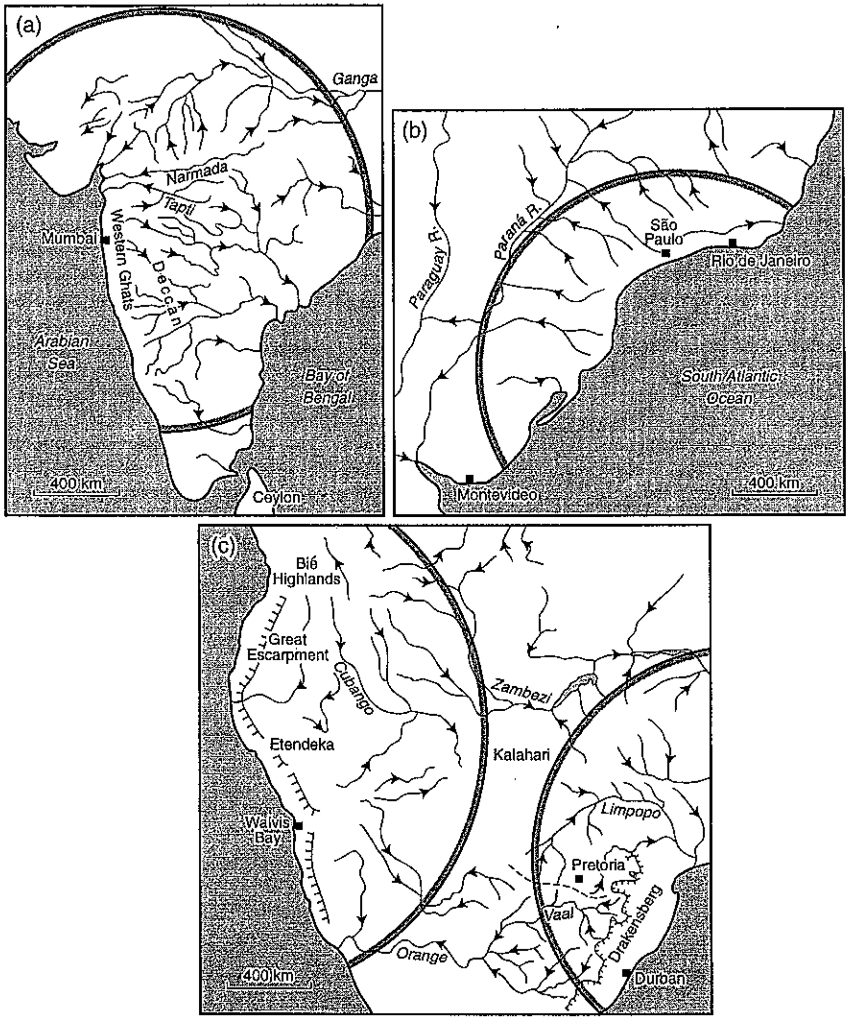

Figure 106 Postulated locations of major domes associated with mantle plumes and showing their relationship to drainage patterns (modified after Cox 1989)

plume origin, and phenomena associated with them include such features as continental flood basalts (e.g. in the Deccan of India), giant dyke and sill swarms (Ernst *et al.* 1995), and large layered intrusions (e.g. the Bushveld Igneous Complex in South Africa). Fourth, mantle plumes may play a major role in the breakup of supercontinents (e.g. Gondwanaland), the formation of passive margins like those in Namibia (Goudie and Eckardt 1999), and the development of basins (e.g. Red Sea, Gulf of Aden, etc.).

References

Condie, K.C. (2001) *Mantle Plumes and their Record in Earth History*, Cambridge: Cambridge University Press.

Cox, K.G. (1989) The role of mantle plumes in the development of continental drainage patterns, *Nature* 342, 873–877.

Ernst, R.E., Head, J.W., Parfitt, E., Grosfils, E. and Wilson, L. (1995) Giant radiating dyke swarms on Earth and Venus, *Earth-Science Reviews* 39, 1–58.

Goudie, A.S. and Eckardt, F. (1999) The evolution of the morphological framework of the Central Namib Desert, Namibia, since the early Cretaceous, *Geografiska Annaler* 81A, 443–458.

Wilson, J.T. (1963) A possible origin of the Hawaiian islands, *Canadian Journal of Physics* 41, 863–868.

A.S. GOUDIE

MASS BALANCE OF GLACIERS

The mass balance of a glacier is the sum of all processes which add mass to a glacier and remove mass from it. Accumulation or addition of mass most commonly takes place in the form of snowfall, often modified by wind and avalanches (see AVALANCHE, SNOW). Melting of snow and ice is the predominant form of ablation or removal of mass, but calving of tidal glaciers, ice avalanching at steep hanging glaciers or evacuation of wind-blown snow in dry areas can locally be of high relative importance. The resulting loss or gain of mass is the direct, undelayed response of a glacier to climatic conditions and must be considered to be a key indicator of climate change (IPCC 2001; Haeberli *et al.* 2002).

Techniques applied to determine glacier mass balance include (a) the direct glaciological method (snowpits and ablation stakes), the geodetic/photogrammetric method (repeated precision mapping), the hydrological method (difference between measured precipitation minus evaporation and runoff) and index methods (snowline mapping, etc.). Long-term mass changes can also be inferred from cumulative glacier length changes using continuity approaches or flow models (Haeberli and Hoelzle 1995; Oerlemans *et al.* 1998).

In humid maritime regions large amounts of ablation are required to compensate for heavy snowfall: the equilibrium line which separates accumulation from ablation areas on glaciers remains at altitudes with relatively warm air temperatures, enabling intense sensible heat flux and strong ice melt during extended ablation seasons. Temperate glaciers at melting temperature, exhibiting high mass turnover and rapid flow, dominate these landscapes. The lower parts of such temperate glaciers commonly extend into grassland and forested valleys where summer warmth and winter snow accumulation prevent the development of PERMAFROST. Ice caps and valley glaciers of Patagonia and Iceland, the western Cordillera of North America and the coastal mountain chains of New Zealand and Norway are features of this type. In contrast, dry continental conditions such as exist in northern Alaska, arctic Canada, subarctic Russia, parts of the Andes near the Atacama desert or in many central Asian mountain chains, force the EQUILIBRIUM LINE OF GLACIERS to elevations with cold air temperatures, short ablation seasons, reduced sensible heat flux and limited amounts of ice melting. In such regions, polythermal or cold glaciers, lying far beyond the treeline and often even beyond tundra vegetation, have a low mass turnover, less rapid flow and are associated with severe periglacial conditions and permafrost (Shumskii 1964).

The goals of long-term mass balance observations are (1) to determine annual ice loss/gain as a regional signal and (2) to better understand processes of energy and mass exchange. A change (δ) in equilibrium line altitude (ELA) induces an immediate change in specific mass balance (b = total mass change divided by glacier area). The resulting change in specific mass balance (δb) is the product of the shift in equilibrium line altitude (δELA) and the gradient of mass balance with altitude (db/dH) as weighed by the distribution of glacier surface area with altitude (hypsometry). The hypsometry represents the local/individual or topographic part of the glacier sensitivity whereas the mass balance gradient mainly reflects the regional or climatic part (Kuhn 1990). As the mass balance gradient tends to increase with increasing humidity and mass turnover (Kuhn 1981), the sensitivity of glacier mass balance with respect to changes in equilibrium line altitude is generally much higher in areas with humid/maritime than with dry/continental climatic conditions (Oerlemans 1993). Rising snowlines and cumulative mass losses lead to changes in average albedo and continued surface lowering. Such effects cause pronounced positive feedbacks with respect to radiative and sensible heat fluxes. In areas of cold firn, atmospheric warming first induces firn warming; mass loss only sets in when the firn has reached melting temperatures and water can leave the system instead of refreezing.

A Global Terrestrial Network for Glaciers as part of the climate-related Global Terrestrial Observing System (GTOS/GCOS) is operated by the World Glacier Monitoring Service (WGMS) which co-ordinates worldwide compilation and dissemination of standardized data on glacier fluctuations. Mass balance measurements are reported in a biennial bulletin (IAHS(ICSI)/UNEP/UNESCO/WMO 2001; http://www.geo.unizh.ch/wgms/). Overall ice loss appears to be strong and probably even accelerating. Anthropogenic greenhouse forcing could have started to exert a predominant influence on this development and may lead to complete deglaciation of many mountain regions of the world within decades (Dyurgerov and Meier 1997a,b; Haeberli *et al.* 2002).

References

Dyurgerov, M.B. and Meier, M.F. (1997a) Mass balance of mountain and subpolar glaciers: a new global assessment for 1961–1990, *Arctic and Alpine Research* 29/4, 379–391.

Dyurgerov, M.B. and Meier, M.F. (1997b) Year-to-year fluctuations of global mass balance of small glaciers and their contribution to sea level, *Arctic and Alpine Research* 29/4, 392–402.

Haeberli, W. and Hoelzle, M. (1995) Application of inventory data for estimating characteristics of and regional climate-change effects on mountain glaciers: a pilot study with the European Alps, *Annals of Glaciology* 21, 206–212. Russian translation: *Data of Glaciological Studies* 82, 116–124.

Haeberli, W., Maisch, M. and Paul, F. (2002) Mountain glaciers in global climate-related observation networks, *WMO Bulletin* 51/1, 18–25.

IAHS(ICSI)/UNEP/UNESCO/WMO (2001) *Glacier Mass Balance Bulletin* No. 6 (Haeberli, W., Hoelzle, M. and Frauenfelder, R. (eds)), World Glacier Monitoring Service, University and ETH Zurich.

IPCC (2001) *Climate Change 2001 – The Scientific Basis*, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge: Cambridge University Press.

Kuhn, M. (1981) Climate and glaciers, *IAHS Publication* 131, 3–20.

——(1990) Energieaustausch Atmosphäre – Schnee und Eis, in *Schnee, Eis und Wasser der Alpen in einer wärmeren Atmosphäre*, Mitteilungen der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie 108, 21–32, ETH Zürich.

Oerlemans, J. (1993) A model for the surface balance of ice masses: part I, alpine glaciers, *Zeitschrift für Gletscherkunde und Glazialgeologie* 27/28, 63–83.

Oerlemans, J., Anderson, B., Hubbard, A., Huybrechts, P., Johannesson, T., Knap, W.H., Schmeits, M., Stroeven, A.P., van de Wal, R.S.W., Wallinga, J. and Zuo, Z. (1998) Modelling the response of glaciers to climate warming, *Climate Dynamics* 14, 267–274.

Shumskii, P.A. (1964) *Principles of Structural Glaciology*, trans. D. Kraus, New York: Dorer.

WILFRIED HAEBERLI

MASS MOVEMENT

A mass movement is the downward and outward movement of slope-forming material under the influence of gravity. The process does not require a transporting medium such as water, air or ice. The term LANDSLIDE often is used as synonymous for mass movement phenomena. However, in a pure sense the term landslide is used as a generic term describing those downward movements of slope-forming material as a result of shear failure occurring along a well-defined shear plane.

There are numerous classifications of mass movements. Most are based on morphology, mechanism, type of material and rate of movement (e.g. Varnes 1978; Hutchinson 1988; Cruden and Varnes 1996). The classification and description used in this entry was developed by a European research group (Dikau *et al.* 1996) (Table 27). The terminology is based upon the classifications of the International Geotechnical Societies' UNESCO Working Party on World Landslide Inventory (WP/WLI) (UNESCO 1993).

Fall

Alternative terms: rockfall, stone fall, pebble fall, boulder fall, debris fall, soil fall.

A fall is a free movement of material from steep slopes. Different types of falls are described by material and failure processes. The term rockfall is often used as the general term without further reference to the material involved. Whalley (1984) and Flageollet and Weber (1996) provide summaries.

Falls occur in various sites such as coastal cliffs, steep riverbanks, edges of plateau, mountain faces or escarpments. They may also occur on artificial embankments (road outcrops). Joints and faults produce planar sheets to form wedge-shaped hollows and boundary vertical joints. Falls can form a fan-shaped cone at the base of the slope. These

Table 27 Classification of slope movements

Type	Rock	Debris	Soil
Fall	Rockfall	Debris fall	Soil fall
Topple	Rock topple	Debris topple	Soil topple
Slide (rotational)	Single (slump) Multiple Successive	Single Multiple Successive	Single Multiple Successive
Slide (translational)	Block slide Rock slide	Block slide Debris slide	Slab slide Mudslide
Lateral spreading	Rock spreading	Debris spread	Soil (debris) spreading
Flow	Rock flow (sackung)	Debris flow	Soil flow
Complex (with runout or change of behaviour downslope, note that nearly all forms develop complex behaviour)	e.g. rock avalanche	e.g. flow slide	e.g. slump-earthflow

Source: Dikau *et al.* (1996)
Note: A compound landslide consists of more than one type e.g. rotational-translational slide. This should be distinguished from a complex landslide where one form of failure develops into a second form of movement i.e. a change of behaviour downslope by the same material

TALUS slopes should be distinguished from accumulations arising from a large fall, a complex rock avalanche (STURZSTROM), which result in an accumulation of debris of all sizes that can block a river, leading to flooding hazards.

Falls are influenced by slope aspect and angle, size and shape of jointed rocks, strike angle, status and deformation of rocks and vegetation cover. Debris and soil falls originate in material which has already been detached from the bedrock. In solid rock the separation process may take time and it arises from internal and external factors which are often combined.

The main implication of fall processes for a planner is to ensure that there are suitable surveys of those areas which are likely to produce rockfalls (Table 27). The main product will be a planning hazard zoning map and suitable monitoring systems. Warning systems are also in operation on various other sites with this type of hazard.

Topple

Alternative terms: rock topple, debris topple, soil topple, tilting blocks.

A topple consists of a forward rotation of a mass of rock, debris or soil about a pivot or hinge on a hillslope. The toppling may culminate in an abrupt falling or sliding, but the form of movement is tilting without collapse. Goodman and Bray (1976) and Dikau *et al.* (1996) provide summaries.

There are several processes responsible for toppling failure such as progressive weathering or erosion resulting in a weakening or loss of elastic underlying material, swelling and shrinking of clay-rich material because of soil moisture changes, and deepening or undercutting of slopes by erosion providing sufficient stress to cause unloading decompression.

The primary driving force for topple failure is the detachment of a column so that the load is transferred to a narrow base of weaker rock. Slope height is an important controlling parameter as is the width of the supporting base. Topples in rocks usually require high cliffs, whereas topples in debris and soil fail on lower cliffs. The formation of tension cracks is caused by severe undercutting by fluvial agents, sea wave action or man-made scarps. Joint and bedding plane water pressures are a vital contribution to failure at the base of the column.

Slide

The term 'slide' is used for a movement of material along a recognizable shear surface. The shear surface type and the number of shear surfaces are used to divide the slide group.

SLIDE (ROTATIONAL)

Alternative terms: slump, rotational slip, rotational slide.

Rotational slides occur as a rotational movement on a circular or spoon-shaped shear surface. They differ in the degree of disintegration in the slide masses and the depositional features in the toe areas. Varnes (1978) and Buma and van Asch (1996) provide summaries.

Varnes (1978) defines a single rotational slide or slump as a 'more or less rotational movement, about an axis that is parallel to the slope contours, involving shear displacement (sliding) along a concavely upward-curving failure surface, which is visible or may reasonably be inferred'. A rotational slide has a small degree of internal deformation although sometimes soil slump material liquefies and transforms into a flow at its toe. In the terminology the American usage 'earthflow' is replaced in European literature by 'mudslide'. Rotational slides can vary from an area of a few square metres to large complexes of several hectares.

Given the relatively small degree of internal deformation, the slump matrix will essentially be the same as the surrounding undisturbed slope. Soil slides generally consist of fine-textured, cohesive materials, like consolidated clays, weathered marls and mudstones. Rotational rock slides often develop in formations of interbedded strong and weaker materials, e.g. marls and limestones or sandstones. In general, rotational sliding produces a disrupted, anomalous drainage pattern. Ponds and peaty areas may develop in depressions between slump units.

Movement of rotational slides starts with initial failure followed by rotation. It may disintegrate into several discrete blocks. In the head area, these blocks tilt backwards while sliding downhill and often flattening or even slope reversal occurs. Sliding along the flanks causes longitudinal and diagonal shear stresses. The lowest part of the slump mass moves over the toe of the failure surface, thereby bulging, cambering, overriding and producing transverse tension cracks.

Movement rates of slumps can vary by several orders of magnitude, between a few centimetres a year to several metres per month, while soil slumps can attain velocities up to 3 metres per second. Tilted trees (generally backwards in the head area, forwards in the foot and toe areas) can reveal the presence of rotational slides.

Typical causal situations are undercutting by waves or streams, excavation and other construction activities. Common triggering mechanisms are earthquakes, explosions and sudden increases of overburden or high water tables following periods of rainfall or snowmelt.

SLIDE (TRANSLATIONAL)

A translational slide is a non-circular failure which involves translational motion on a near planar slip surface. The movement is largely controlled by surfaces of weakness within the structure of the slope-forming material. Translational slides may occur in three types of material: rock, debris and soil. Depending on the slope angle and the velocity, slides will either stay as a discrete block on the failure surface or break into debris.

Block slide

Alternative terms: planar rock slide (rock block slide), slab slide (common usage for soil/earth block slide).

Large block slides are often part of extensive compound landslides involving rotational slides either at the toe or the head and occasionally mudslides at the edges of the landslide. Due to the geometry of the slip surface the mass may only move through the development of internal shears and displacements. Hutchinson *et al.* (1991) and Ibsen *et al.* (1996b) provide summaries.

Block slides are found in stiff, fissured or overconsolidated clays often in combination with stronger rock formations. The basal surface needs only a very low-angle for displacement, because the driving forces are usually very large. Large block slides are particularly sensitive at the toe and sometimes require only slight toe erosion to trigger failure. Movements involve deep settlement at the head, whilst the margins of the block slide are generally totally destroyed.

Block slides can be initiated or reactivated where construction loads the slope or where excavation undercuts or unloads the toe of an area of strongly developed joints or bedding planes that dip outward toward the natural slope. Block slides may move continuously in frequent pulses. In large slides movement is primarily controlled by wet year sequences and extreme or intense rainfall events. Though velocities are frequently low, the masses involved can be very great and extremely difficult or impossible to stabilize.

The primary cause of block slides is the presence of an abrupt change of rock type or a bedded rock sequence which provides a weak stratum dipping gently towards the slope. Strong discontinuities parallel to and into the face which clearly define a potential movement area are also helpful. The second condition concerning block failure is that the slope may be unloaded by erosion or excavation to a point where the potential failure surface crops out above or close to the base level.

Slab slide

Alternative terms: debris block slide, soil block slide, earth block slide, sheet slide (shallow translational failure in dry, cohesionless soils).

Slab slides are translational failures in slopes composed of coherent, fine soils or coarser debris with a fine matrix. Weathered soils, especially those derived from clays, mudrocks and silty-clays, are commonly involved. The weathered material normally moves on a shear zone close to a surface of unweathered or lightly weathered bedrock, a pedogenic horizon or a structural surface. A slab slide is dominated by the pedological or geological structure and frequently fails along discontinuities. Hutchinson (1988) and Ibsen *et al.* (1996a) provide summaries.

Slab slides are extremely susceptible to seasonal changes in groundwater levels or to loading at the head or unloading at the toe. Movement increases in wet months and may cease in dry periods. If the ground freezes and seasonally thaws the saturated conditions can initiate movement. Slab slides also occur at permeable/impermeable soil junctions. Movement takes place on low angle shear surfaces and is normally parallel to the ground surface.

Slab slides are caused by geometrical changes to the slope, water regime changes or human activity. They are dominated by the relationship between regolith depth and slope angles which determines the critical depth for failure. Melting of PERMAFROST is a special cause of saturated ground overlying an impermeable horizon.

Rock slide

A rock slide is a translational movement of rock which occurs along a more or less planar or gently undulating surface (Varnes 1978). It is typical for mountain slopes or rock exposures where the slope angle is close to, or parallel to, the dip of the rock. The movement is controlled by planar structural discontinuities, such as faults, joints and layering and the presence of weaker formations within the rock mass. Sorriso-Valvo and Gulla (1996) and Erismann and Abele (2001) provide summaries.

Rock slides are characterized by well-defined head scarps and flanks, a pronounced scar generally left with little or no debris, and a mass of debris that accumulates in the track or at the base. In case of a sliding rock mass a planar slope surface is developed. If the rock slides well away from the depletion zone, the scar and flanks may remain visible.

There are different mechanisms of movement of rock slides. If the movement is slow (mm to m/day) the whole mass may disaggregate because of differences in velocity along the yield surface. The frequency of events and magnitude of each single slide may vary. In velocity slides the mass disaggregates during movement, transforming it into a rock avalanche or a debris flow.

The essential prerequisites are steep slopes, intense jointing, bedding or fault planes dipping towards the open face and the preparation of the slope by unloading and weathering and the development of joint water pressures. The triggers are the undercutting of the toe support and earthquakes. The fundamental cause of a rock slide is the presence of a rock mass which produces such a stress that the resistance of the intact rock or the friction mobilized on existing discontinuities is exceeded.

Rock slides have a wide range in volume and velocities and pose considerable hazards to human settlements and lives. The destructive power of rock slides can be enormous in the case of rapid rock slides on steep slopes.

Debris slide

Alternative terms: shallow translational slides, sheet slides, soil slips.

Debris slides are failures of unconsolidated material which breaks up into smaller parts as the slide advances downslope. The material involved is mostly COLLUVIUM and weathered material of fractured rocks masses (i.e. flysch formations, shales and slates). The failure surface usually develops at the contact between the REGOLITH cover and the bedrock, and is roughly parallel to the ground surface. Clark (1987) and Corominas *et al.* (1996) provide summaries.

The speed of sliding and degree of runout tend to increase with slope angle and decrease with clay content (Hutchinson 1988). Velocities of up to 16 m/sec have been recorded. Many translational debris slides turn into debris flows. This occurs where water is available, and where the topography favours the convergence of both debris and water into concavities and channels. On very steep slopes, debris slides can reach high velocities. This is common in valleys shaped by glaciers in which morainic sediment is located high above the present river.

Debris slides are often triggered by intense rainfall or by earthquakes. The probability of a debris slide occurring is greatly increased by the destruction of vegetation cover by fires or logging. Sites most likely to provide failures are first-order basins with hollows where regolith can reach the maximum thickness and high slope angles. Failures are often caused by an increase in PORE-WATER PRESSURES following heavy rains which reduce the shear strength of the material. After failure, the breakage of the sliding mass allows the water to escape and the debris to stop. Debris slides can burst explosively out of a slope. Hazard assessments should be based on regolith depth and slope angle.

Mudslide

Alternative terms: earthflow (US usage), mudflow (redundant usage also climatic and temperate mudflow), slump-earthflow (complex, lobate mudslide form with minimal track).

Mudslides are a form of mass movement in which masses of softened silty or very fine sandy debris slide on discrete boundary shear surfaces in relatively slow-moving lobate or elongate forms. Brunsden (1984) and Ibsen and Brunsden (1996) provide summaries.

A mudslide is divided into source, track and lobe and accumulation zone units. The source has a bowl-shaped head. The material in this section is usually soft, weathered debris often with depressions containing water. The track evolves as an elongate or lobate channel through which the material moves. The accumulation zone develops at the base of the slope and normally consists of lobes of debris. Mudslides usually occur in saturated clays which have been described as fissured, mudstones, siltstones and overconsolidated clays with a generally medium plasticity.

Mudslide movement rates range from about 1–25 m/yr and are generally classified as slow mass-movement types. Extreme events range from hundreds of metres per day. Mudslide movement is normally seasonal, as the wetter weather increases the water content to the point where pore water becomes sufficient to generate movement. Mudslides in temperate areas display a pronounced winter–summer cycle. Movement usually commences in the late autumn, peaks in mid-winter and comes to a slow halt by late spring and summer. Heavy rainfall will frequently result in a mudslide surge. Movement can also be generated by undrained loading when there is rapid supply of debris at the head. Other causes are associated with the supply of water such as snowmelt or permafrost melt. Finally unloading of the toe area is important since it presents the development of passive resistance at the toe.

The essential plannning and engineering implication is that this form of slide requires water. Planning must include surveys for drainage installations. The removal of the source of water influx is critical, as is the reduction of pore-water pressure from the mudslide mass. Roads and other linear features are most vulnerable to movement at the lateral shears.

Lateral spreading

Lateral spreading describes the lateral extension of a cohesive rock or soil mass over a deforming mass of softer underlying material.

ROCK SPREADING

Alternative terms: gravitational spreading, gravity faulting, block-type slope movement, cambering and valley bulging.

Rock spreading is the result of deep-seated, plastic deformation in the rock mass, leading to extension at the surface. It may take place in a relatively homogeneous rock, or there might be a fractured capping stratum leading to gravitational stresses. Where the rock is relatively homogeneous, the moving mass breaks up into successive units arranged like horsts and grabens. Rock spreading in homogeneous rock masses is characterized by double ridges, trenches and uphill facing scarps which occurs often in high mountains. Uphill facing scarps evolve from the combination of ridge spreading and erosion of the uphill sides of the tension cracks. Zaruba and Mencl (1982) and Pasuto and Soldati (1996) provide summaries.

Rock spreading is very often associated with toppling, rock falls, slumps and mudslides. The rock masses involved are huge and generally more than one million cubic metres. The velocity is particularly low in comparison with other types of mass movement and seasonality is rarely observed. Lateral spreading phenomena are highly controlled by geological structures and often connected with deep-seated gravitational slope deformations.

Rock spreading is caused by an outward and downward movement of both valley sides along low-angle shear planes causing rock spreading at the ridge top. There may be sliding of blocks with respect to the whole ridge with minor crushing and redistribution of material at the point of the wedges. Alternative wedging along the centre of the ridge in combination with rotation of the peripheral blocks together with crushing at the areas of greatest stress may take place. Rock spreading rates range between 10^{-4} and 10^{-1} metres per year. Large-scale rock spreading is generally not seasonal.

The process depends on the local geological and topographical conditions. The horizontal or subhorizontal overposition of a thick slab of competent rocks, affected by a dense network of tectonic joints, on clayey materials which may behave as a visco-plastic medium, is a prerequisite for the process.

SOIL/DEBRIS SPREADING

Alternative terms: sudden spreading failure, lateral soil spreading, quick clay sliding, quick clay flow, soil liquefaction sliding.

Soil spreading is defined as the collapse of a sensitive soil layer followed by either settlement of the overlying more resistant soil layers, or progressive failure throughout the whole sliding mass. The duration is generally only a few minutes. Considerable lateral movement along the basal mobile zone occurs. The material deformations accompanying the spreading often cause loss of life and severe damage to roads, buildings and embankments. The areas involved are low angled and are ideal for both agricultural use and urban development. Areas of particularly intense hazards are often situated close to the shoreline or the banks of rivers, as toe erosion is a significant factor in destabilizing these slopes. Bjerrum (1955) and Buma and van Asch (1996) provide summaries.

QUICKCLAYS are found in areas which have been depositional marine environments adjacent to Pleistocene ice margins. Morphological features of quickclay slides which have their whole slide mass liquefied are the pear-shaped scar, and the presence of an extensive flow lobe on the lower side of the scar. This lobe can attain lengths of up to 1,000 m. Quickclay slides start in lower slope regions and extend upslope by retrogression. Due to loss of horizontal support, retrogression proceeds rapidly, both upslope and to the flanks. Due to remoulding the clay already subject to movement loses its internal strength and descends the slope as a more or less viscous slurry, and can travel up to several kilometres. In (quick) clay soils, initial failures are mostly caused by extremely high water contents (30–40 per cent) within the clay, oversteepening of the local slope and/or stress caused by loading.

Quickclay slides are initiated within a body of sensitive ('quick') clay. Sensitive clays are defined as clays which need only very little remoulding to transform into a viscous slurry. However, the whole sliding mass does not necessarily liquefy. It may be restricted to a thin layer of sensitive clay at some depth. The sensitivity of the clay is influenced by the dominant type of clay minerals and the depositional environment. Marine sedimentation results in flocculation and an open 'card-house' porosity structure. Strength reduction through interparticle repulsion is caused by leaching of salt in pore water over time.

Sand liquefaction slides are situated on coasts and adjacent to rivers or lakes. Especially sensitive are slopes already subject to mass movement, with disturbed drainage and corresponding high water contents in the soil layers. Varved clay formations are also sensitive to sand liquefaction failures along the boundary between impermeable clay layers and water-bearing fine sand or silt layers under high pore-water pressures. Tension fractures will develop at the head of the slide, dipping towards the sliding mass. The overhanging wall of the sliding block collapses to form a graben. The sliding blocks become subject to tilting, internal fracturing, subsidence, heaving and overthrusting, producing a very hummocky topography. The liquefied layer material escapes through tension cracks and may cause sand boils.

A triggering mechanism of sand liquefaction is a combination of prolonged periods of heavy rainfall or snowmelt, causing high initial pore pressures, and earthquakes. During an earthquake a soil undergoes cyclic stress loading and unloading, and pore-water pressures increase with each cycle. After a certain number of cycles the pore-water pressures become equal to the existent confining pressures and the soil suddenly loses its strength, suffering from considerable deformations without exhibiting resistance. This process is restricted to cohesionless soils such as sand and silt.

Flow

A flow is a landslide in which the individual particles travel separately within a moving mass. They involve highly fractured rock, clastic debris in a fine matrix or small grain sizes. Flow in its physical sense is defined as the continuous, irreversible deformation of a material that occurs in response to applied stress. They are, therefore, characterized by internal differential movements that are distributed within the mass.

ROCK FLOW

Alternative terms are: SACKUNG, sagging, rock creep, deep-seated gravitational creep.

Rock flows are creeping flow-type, deep-seated gravitational deformations affecting homogeneous rock masses. Rock flows are characterized by the high volume of the rock mass (several thousand to millions of cubic metres) and small total displacement rates. Structural elements and landforms associated with rock flows are high angle extensional shear planes in the upper part of the deforming slopes, producing graben-like depressions (trenches), double ridges, ridge depressions and troughs. The slope foot often shows compressional features such as bulging and, sometimes, low-angle shear planes. Mencl (1968) and Bisci *et al.* (1996) provide summaries.

The mechanisms of rock flows are not well known. Mencl (1968) postulates that in the central part of the slope, where the confining pressures are high and the deviator stress is too small to produce shearing, the rock mass deforms through viscous flow. At the uppermost and lowermost parts of the slope, where the confining pressures are low, the rock mass moves along the shearing surfaces. At depth, the high pressures induce plastic deformation without necessarily creating a proper slip surface.

Some rock flows display evidence of a constant rate of creep deformation. Others show a rapid reactivation phase in connection with extreme rainfall or earthquakes. Some rock flows may be characterized by a step-like evolutionary behaviour, including short active phases related to critical events and long-lasting dormant phases involving creep deformation at extremely slow rates. Under particular circumstances, the slow slope deformation can be turned into a catastrophic event leading to the collapse of a large-scale rock avalanche or debris flow. Rock flows can be considered as extremely slow preparatory stages of huge landslides which only in a few cases reach their final evolutionary step.

Rock flows are produced only where slopes are high enough to induce strong gravitational stress in the bedrock. Such conditions are typical of valley slopes in mountain areas and high coastal cliffs. Rock flows do not normally represent a major problem to planning and engineering. In recent years, however, there is a concern that large structures such as dams or hydrological sites may be placed at the toe of these phenomena.

DEBRIS FLOW

Alternative terms: mudflow (old usage), lahar (volcanic mudflow).

Debris flows consist of a mixture of fine material (sand, silt and clay), coarse material (gravel and boulders), with a variable quantity of water, that forms a muddy slurry which moves downslope. The flow moves in surges and includes the erosion of the channel bed and the collapse of bank material. Debris flows usually take place on slopes covered by unconsolidated rock and soil debris. Debris flows are characterized by the source area, the main track and the depositional toe. The flows commonly follow existing drainage ways. Some of the coarse debris will be deposited at the side of the track to form lateral ridges (levees). Deposits are accumulated where the channel gradient decreases or at the toe of mountain fronts. Successive surges will build up a debris fan. Some debris flows are exceptionally large, fluid, and can reach long distances beyond the source area. Debris flow is a gravity induced mass movement between landsliding and water flooding, although with mechanical characteristics very different from either of these processes. Pierson and Costa (1987) and Corominas *et al.* (1996) provide summaries.

Debris flows are a very destructive type of mass movement caused by heavy rain or snowmelt. In alpine environments debris flows are composed of coarser material from mechanical weathering and glacial deposits. Debris material in melting permafrost (near the lower limit of the discontinuous permafrost) has also been considered a debris flow source area. Debris flows which originate on the slope of a volcano contain vulcaniclastic materials and are called LAHARS.

Debris flows consist of large coarse material embedded in a fine-grained matrix. The coarse material is randomly distributed and individual beds are generally poorly sorted. Buoyant forces and dispersive pressures may concentrate boulders at the top of the deposit, forming reverse grading. The frequency of debris flow events is controlled by the rate of accumulation in hollows or channels, and by the recurrence of climatic triggering events.

Debris flows have well-graded deposits with a small clay content, generally less than about 5 per cent. They have a range in volume concentration of solids from approximately 25 to 86 per cent. Sediment concentration is the primary criteria for classification of flows given by Pierson and Costa (1987). There is a continuum from sediment movement in rivers to debris flows. Fluids with large sediment concentrations do not deform until a threshold strength is exceeded and they behave like a non-Newtonian fluid.

Debris flows are commonly triggered by an unusual presence of water created by intense rainfall, rapid snowmelt, and glacier or lake overflows. Debris flows are frequent in topographic concavities or hollows at first-order watersheds. This geometry supports the accumulation of colluvium and the convergence of groundwater flow necessary to cause the failure. Many debris flows start as a translational or rotational slide and then turn into a debris flow.

Series of debris flow waves are frequent and can be produced by breaching of the temporary dams or obstructions in the channel. The front lobe is composed of coarse blocks occasionally mixed with trees. During the progression of the debris flow through a channel, lateral ridges (levees) can develop. Overflowing of channel banks is a significant natural hazard of debris flows. The velocity of the flow depends on the size and sediment concentration, and on the geometry of the path. Observed velocities range from 0.5 to about 20 m/s. Large lahars can travel over a distance of more than 100 km and the rate of movement may reach more than 50 km/h. The erosion that occurs on both the channel floor and banks, cause some debris flows to significantly increase their volume.

The socio-economic impact and the loss of life, property and agriculture can be catastrophic. Even smaller mudflows and debris flows may cause serious damage, especially in mountainous regions. The deposits are also responsible for severe indirect damage and hazards such as damming of rivers or sudden debris supply to river systems. It is essential that potential source areas and runout zones are assessed.

SOIL FLOW (MUDFLOW)

Alternative terms: mudflow, alpine mudflow, earthflow, sandflow.

Soil flows may occur in three forms, wet mudflow, wet sand flow and dry sand flow. The wet forms are special categories of debris flow where the material is of a single and fine grain size and coarse clasts are rare. They are very mobile and can flow downslope quickly. They tend to follow gullies or shallow depressions and then to spread out into a flat fan or even a thin sheet when they reach low gradients. Soil flow conditions are abundant water, unconsolidated material and insufficient protection of the ground (i.e. lack of vegetation). Pierson and Costa (1987) and Schrott *et al.* (1996) provide summaries.

A wet soil flow contains relatively cohesive earth material that has at least 50 per cent sand, silt and clay. Thus, the term soil flow should be used for a flow with a significant lack of coarse-grained material. Typical source areas and starting zones are steep (25°–40°) slopes (e.g. moraines, proglacial zones), volcanic environments and mountain torrents.

A characteristic of soil flows is their ability to travel long distances (some kilometres) over even low slopes, usually following pre-existing drainage patterns. They are often termed viscous slurry flows, but they can be either viscous or inertial flows, depending on the driving forces. The flow behaviour is normally that of a visco-plastic type. The change from a slow creep to visco-plastic flow in clay-rich soils supersedes the destruction of strong bonding and the subsequent decrease in viscosity. Continuous undrained loading causes rapid readjustment of this mass and velocities of up to 10 m per second have been recorded.

Very rarely dry sand flows may occur. These form when a large mass of dry material falls from or over a steep slope and fluidizes on impact. Flow is then of 'rock avalanche' type with a track of uniform depth. Dry sand flows may also occur from dunes or similar sandy deposits. They require the sliding of material at the head and then either the transfer of momentum by cohesionless grain flow (i.e. grain upon grain momentum transfer) or descent over a cliff and fluidization. The first type can be observed in fluviglacial deposits or riverbanks.

Complex

It is common for mass movement processes to combine together and complex landslides occur where the initial failure type changes into another as it moves downslope. Compound landslides include two types of movement which occur concurrently within the same failure.

ROCK AVALANCHE

Alternative terms: STURZSTROM, rockfall avalanche, rock-slide avalanche.

FLOW SLIDE

A flow slide is a structural collapse of slope-forming material with momentary fluidization and is usually referred to as a high magnitude event both in terms of velocity and destruction. The high energy is capable of causing incredible devastation not merely through its impact on humans, by cutting communication and power lines or diverting a river, but also environmentally by littering the surrounding valley with debris. Flow slides are often associated with man-made tips and spoil heaps, although this type of failure may also occur in rock debris of geological origin. The internal structure has very little cohesion and the matrix ranges from clay-size particles to large blocks. Flow slides are a subclass of debris flows. The various causes of flow slides are an initial rotational failure at the head, vibrations or shocks, heavy rainfall, loosely deposited spoil heaps, removal of lateral support and rapid loading. Bishop *et al.* (1969) and Ibsen *et al.* (1996c) provide summaries.

The key characteristics of a flow can lie in their origin in artificial spoil materials but their behaviour is also found in many natural debris flow events. A flow slide is generally composed of loose material, which losses its cohesion with a reduction in strength, becoming a fluidized mass. The fluid may be air or water and, therefore, the dominant mechanism of a flow slide is fluidization or liquefaction. Sliding may occur at the head of the flow slide perhaps in the form of a rotational slide, but there is usually little indication of shearing at the subsequent stages of movement. Flow slides not only fluidize very quickly, but also rapidly consolidate and become solid when they cease moving, creating an additional hazard in the depositional area.

References

Bisci, C., Dramis, F. and Sorriso-Valvo, M. (1996) Rock flow (sackung), in R. Dikau *et al.* (eds) *Landslide Recognition*, 150–160, Chichester: Wiley.

Bishop, A.W., Hutchinson, J.N., Penman, A.D. and Evans, H.E. (1969) *Geotechnical investigations into the causes and circumstances of the disaster of 21st October 1966. A selection of technical reports submitted to the Aberfan tribunal*, Welsh Office, London: HMSO.

Bjerrum, L. (1955) Stability of natural slopes in quick clay, *Géotechnique* 5(1), 101–119.

Brunsden, D. (1984) Mudslides, in D. Brunsden and D.B. Prior (eds) *Slope Instability*, 363–418, Chichester: Wiley.

Buma, J. and van Asch, T. (1996) Slide (rotational), in R. Dikau, *et al.* (eds) *Landslide Recognition*, 43–61, Chichester: Wiley.

Clark, G.M. (1987) Debris slide and debris flow historical events in the Appalachians south of the glacial border, *Review of Engineering Geology* 7, 125–138.

Corominas, J., Remondo, J., Farias, P., Estevao, M., Zezere, J., Diaz de Teran, J., Dikau, R., Schrott, L., Moya, J. and Gonzalez, A. (1996) Debris flow, in R. Dikau *et al.* (eds) *Landslide Recognition*, 161–180, 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.

Dikau, R., Brunsden, D., Schrott, L. and Ibsen, M.-L. (1996) *Landslide Recognition*, Chichester: Wiley.

Dikau, R., Schrott, L. and Dehn, M. (1996) Topple, in R. Dikau *et al.* (eds) *Landslide Recognition*, 29–41, Chichester: Wiley.

Erismann, T.H. and Abele, G. (2001) *Dynamics of Rockslides and Rockfalls*, Heidelberg: Springer.

Flageollet, J.C. and Weber, D. (1996) Fall, in R. Dikau *et al.* (eds) *Landslide Recognition*, 13–28, Chichester: Wiley.

Goodman, R.E. and Bray, J.W. (1976) *Toppling of Rock Slopes*, Proceedings of Special Conference on Rock Engineering for Foundation 2, 201–234.

Hutchinson, J.N. (1988) Morphological and geotechnical parameters of landslides in relation to geology and hydrology, General Report, in C. Bonnard (ed.) *Landslides*, Proceedings of the 5th International Symposium on Landslides 1, 3–35.

Hutchinson, J.N., Bromhead, E.N. and Chandler, M.P. (1991) Investigations of landslides at St. Catherine's Point, Isle of Wight, in R.J. Chandler (ed.) *Slope Stability Engineering Developments and Applications*, Proceedings of the International Conference of Civil Engineers, Isle of Wight, 169–179.

Ibsen, M.-L. and Brunsden, D. (1996) Mudslide, in R. Dikau *et al.* (eds) *Landslide Recognition*, 103–119, Chichester: Wiley.

Ibsen, M.-L., Brunsden, D., Bromhead, E. and Collision, A. (1996a) Slab slide, in R. Dikau *et al.* (eds) *Landslide Recognition*, 78–84, Chichester: Wiley.

Ibsen, M.-L., Brunsden, D., Bromhead, E. and Collision, A. (1996b) Block slide, in R. Dikau *et al.* (eds) *Landslide Recognition*, 64–77, Chichester: Wiley.

Ibsen, M.-L., Brunsden, D., Bromhead, E. and Collision, A. (1996c) Flow slide, in R. Dikau *et al.* (eds) *Landslide Recognition*, 202–211, Chichester: Wiley.

Mencl, V. (1968) Plastizitätslehre und das wirkliche Verhalten von Gebirgsmassen, in *Felsmech. und Ing. Geologie*, Suppl. 4, 1–8.

Pasuto, A. and Soldati, M. (1996) Rock spreading, in R. Dikau *et al.* (eds) *Landslide Recognition*, 122–136, Chichester: Wiley.

Pierson, T.C. and Costa, J.E. (1987) A rheologic classification of subaerial sediment-water flows, Geological Society of America, *Reviews in Engineering Geology* 7, 1–12.

Schrott, L., Dikau, R. and Brunsden, D. (1996) Soil flow (mudflow), in R. Dikau *et al.* (eds) *Landslide Recognition*, 181–187, Chichester: Wiley.

Sorriso-Valvo, M. and Gulla, G. (1996) Rock slide, in R. Dikau *et al.* (eds) *Landslide Recognition*, 85–96, Chichester: Wiley.

UNESCO (1993) *Multilingual Landslide Glossary*, The International Geotechnical Societies' UNESCO Working Party for World Landslide Inventory, BiTech Pubs, Canada.

Varnes, D.J. (1978) Slope movement types and processes, in R.L. Schuster and R.J. Krizek (eds) *Landslides analysis and control*, Transportation Research Board, National Academy of Sciences, Special Report 176, 12–33.

Whalley, W.B. (1984) Rockfalls, in D. Brunsden and D.B. Prior (eds) *Slope Instability*, 217–256, Chichester: Wiley.

Zaruba, Q. and Mencl, V. (1982) Slope movements caused by squeezing out of soft rocks, in Q. Zaruba and V. Mencl (eds) *Landslides and their Control*, 110–120, Amsterdam: Elsevier.

Further reading

Bromhead, E.N. (1992) *The Stability of Slopes*, 2nd edition, New York: Surrey University Press/Chapman and Hall.

Brunsden, D. (1993) Mass movement; the research frontier and beyond: a geomorphological approach, *Geomorphology* 7, 85–128.

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

SEE ALSO: factor of safety; fluidization; lahar; landslide; landslide dam; method of slices; quickclay; riedel shear; sackung; sensitive clay; sturzstrom

RICHARD DIKAU

MATHEMATICS

Mathematics provides an essential set of tools for the study of geomorphological phenomena. Aspects of mathematics particularly relevant to geomorphology include analysis of numerical relationships among measured quantities, the use of differential equations to describe processes, and descriptive mathematics related to form, such as geometry and topology. The former two aspects are critical when deriving equations to describe geomorphological processes or statistical relations between system variables. The latter is relevant when considering morphology of landscape features.

Prior to the twentieth century, the study of Earth sciences in the western world consisted mainly of qualitative consideration of early Earth history and landform origins. Until the mid-twentieth century, mathematics in geomorphology was limited to a few pioneering studies or engineering problems. Davis (1899), following earlier traditions, influenced geomorphological research for many decades by adopting a qualitative framework to describe regional landscape evolution in his 'CYCLE OF EROSION'. At around this time, Gilbert (1877) developed an innovative, quantitative research agenda based on the application of Newtonian mechanics to the study of landscape processes. This latter approach was not generally adopted at the time, perhaps because of difficulties in measuring processes at the large scales of study that dominated the discipline.

Around the mid-twentieth century, the 'quantitative revolution' in geomorphology emerged. Landmark papers such as Horton (1945) and Strahler (1952) signalled the shift to a quantitative approach. Morphometrics and numerical analysis of processes became dominant research themes. Scales of study decreased, as it is generally easier to measure system parameters at smaller scales. In the ensuing decades, many studies focused on establishing empirical relations between system variables. Physically meaningful expressions, in contrast to empirical relations, must be dimensionally balanced. In fact, what may appear to be fundamental relations in geomorphology do not actually fit this definition. Such equations, therefore, represent scale relations and are not true representations of the underlying physics. Physically based attempts to describe processes make use of mathematical tools, such as differential calculus, to portray fundamental conservation principles (mass, momentum, energy) in temporally evolving geomorphological systems. The Buckingham Pi Theorem, which is based on formal dimensional analysis, can be used in an attempt to derive rational equations that encompass essential system attributes.

The use of mathematical techniques is now firmly entrenched in geomorphological research. Continuing advances in technologies, such as remote sensing and radiometric dating, and the availability of improved data sets, such as DIGITAL ELEVATION MODELS (DEMs), have made it possible to measure and describe geomorphic phenomena at large spatial and temporal scales. Although complexities in the natural environment prohibit exact quantification of geomorphic processes, simplified numerical models of landscape processes have emerged due to advances in computing technologies. Sensitivity analysis, which examines the significance of changes in controlling variables on process operation, can be undertaken within a modelling framework.

References

Davis, W.M. (1899) The Geographical Cycle, *Geographical Journal* 14, 481–504.
Gilbert, G.K. (1877) *Report on the Geology of the Henry Mountains*, US Geographical and Geological Survey of the Rocky Mountain Region, Washington, DC.
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.
Strahler, A.N. (1952) Dynamic basis of geomorphology, *Geological Society of America Bulletin* 63, 923–938.

YVONNE MARTIN

MAXIMUM FLOW EFFICIENCY

The operation of fluvial systems can be described by the physical relationships of flow continuity, resistance and sediment transport. However, these three relationships are applicable to any alluvial-channel section and their solution involves more than three variables (width, depth, velocity and slope). As a result, the number of sections satisfying the three relationships is generally very large. In an attempt to understand the physics behind this problem of non-closure solution, Huang and Nanson (2000, 2002) propose a mathematical analytical approach that identifies a mechanism of self-adjustment. Among the numerous channel sections that satisfy flow continuity, resistance and sediment transport, a unique solution of HYDRAULIC GEOMETRY occurs when flow reaches the following inherent, optimal state defined as maximum flow efficiency (MFE):

$$\varepsilon = \frac{Q_s}{\Omega^\lambda} = a\ maximum$$

where ε is flow efficiency factor, Q_s is sediment discharge, Ω is stream power or $\Omega = \rho g Q S$, and exponent λ has a value of 0.65–0.85. For two reasons this state can be regarded as a fundamental law governing the adjustment of fluvial systems. First, MFE can be directly derived from the widely applied LEAST ACTION PRINCIPLE (LAP) (Huang *et al.* 2002). Second, mathematically derived MFE channel geometry relations are highly consistent with empirical relations developed from a wide range of observations for stable canals and relatively stable river channels (Huang and Nanson 2000, 2002).

While recognition of the applicability of LAP to river systems in the form of MFE provides a soundly based and computationally economical method for determining stable alluvial channel geometry, its physical implications are much more profound. First, it is an advance over the thermodynamic analogies and the empirical formulas previously used to justify numerous extremal MODELS in geomorphology, for it clarifies confusion regarding which of these hypotheses should be regarded as rationally based. Indeed, MFE subsumes the earlier hypotheses of maximum sediment transport capacity and minimum stream power. Second, it identifies a fundamental cause for the formation of different river channel patterns. It is the balance between available stream power and imposed sediment load that ultimately determines equilibrium channel form. This balance has been shown to exist in the ideal case of a single-thread, straight and fully adjustable channel system (Huang and Nanson 2000, 2002). Ongoing research suggests that when the balance in the ideal system cannot be maintained due to the effect of physical restrictions, such as the imposed valley gradient, then planform or cross-sectional changes will occur to either consume excess energy or to increase transport efficiency over low gradients. As a consequence, meandering, braiding, anabranching and wandering channel patterns will be formed (Huang and Nanson 2002).

References

Huang, H.Q. and Nanson, G.C. (2000) Hydraulic geometry and maximum flow efficiency as products of the principle of least action, *Earth Surface Processes and Landforms* 25, 1–16.
Huang, H.Q. and Nanson, G.C. (2002) A stability criterion inherent in laws governing alluvial channel flow, *Earth Surface Processes and Landforms* 27, 929–944.
Huang, H.Q., Nanson, G.C. and Fagan, S.D. (2002) Hydraulic geometry of straight alluvial channels and the principle of least action, *Journal of Hydraulic Research* 40, 153–160.

SEE ALSO: hydraulic geometry; least action principle; models

HE QING HUANG AND GERALD C. NANSON

MEANDERING

Phenomenology of meandering

Meandering refers to the spontaneous evolution of a single channel to high values of SINUOSITY, or to a channel that shows this pattern (Plate 78). Meandering is one of three basic types of river planform, of which the other two are straight and braided (see BRAIDED RIVER) (or, more generally, anabranching). Individual channels of anastomosed rivers can show meandering (see ANABRANCHING AND ANASTOMOSING RIVER). Meandering with planform shape comparable to that in rivers also occurs in tidal channels, in marine channels produced by density CURRENTS, and in geostrophic ocean-surface currents like the Gulf Stream.

Plate 78 The meandering Pembina River, Alberta, Canada. Flow towards top of image

In classical meandering the width of channel remains constant as the sinuosity increases, so the channel planform is well described by a single sinuous line. The simplest function that gives an adequate description of this curving line is a relative of the common sine wave known as a sine-generated curve. A sine-generated curve is one in which the local direction of flow varies as the sine of distance along the channel. As meandering reaches high sinuosity values, the symmetric sine-generated curve develops an asymmetry that gives the channel path a more saw-toothed shape (Parker and Andrews 1986). The sharp 'corner' of the tooth is on the upstream side of the bend.

Origin of meandering

Observationally, meandering is favoured in fine-grained (sand and finer) rivers with high suspended load relative to bedload and low slopes, but it is known from steep and/or coarse-bedded rivers as well, so the former conditions are apparently not fundamental. Meandering river channels are also often associated with fine-grained and/or vegetated banks and usually have well-developed floodplains.

The first step towards understanding the mechanics of meandering was to understand the effects of streamline curvature on flow in channel bends. Rozovskii was the first to show mechanistically how streamline curvature leads to a secondary flow, that is, a closed circulation of the water across the direction of the vertically averaged flow. The secondary circulation moves fast-moving surface water toward the outer bank and slowly moving bottom water toward the inner bank. The net effect of this circulation pattern is erosion on the outer bank and deposition on the inner bank, leading to development of a bank-attached bar (a point bar) on the inner bank. Thus channel curvature tends to be self-amplifying – an example of positive feedback. As the bend grows, deposition on the inner bank often maintains a rough balance with erosion of the outer bank, keeping the width constant. The growth of the point bar is often recorded in the surface morphology as a set of scroll bars that trace previous positions of the inner bank (Plate 78).

Later work has elaborated on this simple model considerably. First, the dynamics of bend flow is influenced at least as much by mean-flow inertia and by bottom topography as by secondary circulation. Interaction of the secondary circulation and the mean flow also tends to displace the thread of maximum velocity downstream relative to the bend apices, leading to a tendency of the bends to migrate downstream.

A major next step was formal mathematical analysis of the stability of a straight channel, with the aim of providing a mechanistic basis for the occurrence of straight, meandering and braided channel patterns (Fredsoe 1978; Parker 1976) (see BRAIDED RIVER). Meandering is thought to correspond to cases where the main mode of instability is alternate bars, which are bank-attached bars on alternating sides of the channel. Alternate bars are predicted to develop in channels with widths of roughly 15 to 150 times the flow depth. The bars alternately deflect the flow from one bank to the other, producing a sinuous thalweg (planform trace of the deepest part of the channel). This initial sinuosity leads to fully developed meandering via the positive-feedback mechanisms discussed above. A more elaborate stability analysis (Blondeaux and Seminara 1985) shows that the meandering instability is actually a kind of 'resonance' between the original alternate-bar instability and a planform instability of the channel curvature. Given that they apply strictly only to the initial growth of the bend, it is remarkable that these stability analyses correctly predict the wavelength of fully developed meanders: about seven times the channel width.

Stability analysis suggests that the main control on channel pattern is the aspect ratio of the channel. Although the dynamics of channel width are still not entirely understood, it is clear that one of the main controls is the total effective sediment flux that must pass through the channel. Apart from directly controlling the width, the effective sediment flux also influences the aspect ratio indirectly: the slope is proportional to the ratio of sediment discharge to water discharge; as slope decreases, depth tends to increase. Hence reducing the effective sediment flux relative to water discharge reduces the width and increases the depth. The central role of the sediment/water discharge ratio in controlling the plan pattern is consistent with the empirical observation that meandering tends to occur in low-gradient rivers with relatively fine, suspension-dominated sediment flux. However, neither the gradient nor the sediment grain size *per se* is the real controlling factor.

Stability analysis helps explain the origin of meandering. However, since it treats only the initial instability, it does not constrain the final amplitude of the fully developed meanders. The ratio of amplitude to meander wavelength sets the sinuosity of the channel. There is still no complete analysis of the controls on the amplitude of fully developed river meanders. The main process that limits meander growth in rivers is cutoff of the bend by formation of a new, shorter channel across the bend. Thus one could view meander geometry and average sinuosity as the outcome of a competition between sinuosity production by meander growth and sinuosity destruction by cutoff (Howard 1992). The ease with which cutoff occurs appears to be controlled by the erodibility and resistance to flow of the point bar surface. Vegetation on the point bar helps prevent cutoff in several ways. Stems and leaves block flow and provide baffling that aids in the deposition of fine, cohesive sediment, while roots bind deposited sediment. It would be helpful if techniques could be developed to reproduce steady-state, fully developed meandering experimentally, but so far this has proved extremely difficult to do (Smith 1998).

Submarine meandering

Submarine channel systems formed by density currents produce meander patterns that are very similar to river meanders in planform geometry. However, the scale of submarine channels is much larger than that of river channels: depths are typically 100–200 m and widths typically several kilometres. The mechanics of submarine meanders are similar to those of subaerial meanders; the larger scale results mainly from the fact that the density difference between turbid and clear water is much less than that between water and air. Hence much deeper flows are required to provide sufficient force to move sediment particles.

References

Blondeaux, P. and Seminara, G. (1985) A unified bar-bend theory of river meanders, *Journal of Fluid Mechanics* 157, 449–470.

Fredsoe, J. (1978) Meandering and braiding of rivers, *Journal of Fluid Mechanics* 84, 609–624.

Howard, A.D. (1992) Modeling channel migration and floodplain sedimentation in meandering streams, in P.A. Carling and G.E. Petts (eds), *Lowland Floodplain Rivers: Geomorphological Perspectives*, 1–41, New York: Wiley.

Parker, G. (1976) On the cause and characteristic scales of meandering and braiding in rivers, *Journal of Fluid of Mechanics* 76, 457–480.

Parker, G. and Andrews, E.D. (1986) On the time development of meander bends, *Journal of Fluid Mechanics* 162, 139–156.

Smith, C.E. (1998) Modeling high-sinuousity meanders in a small flume, *Geomorphology* 25, 19–30.

CHRIS PAOLA

MECHANICAL WEATHERING

Mechanical weathering causes disintegration of rock, without substantial chemical and mineralogical alteration or decomposition. The culmination of mechanical weathering is the collapse of parent material and diminution of clast size. Rock disintegration is caused by stresses exerted along zones of weakness in the material, which may include pre-existing fractures, bedding planes or intergranular boundaries.

Mechanical weathering processes promote rock breakdown by inducing stresses within the rock; these stresses may be produced by volumetric change in the rock itself or by deposition of, and/or volumetric change in, material introduced into voids in the rock. Volumetric expansion of rock may be induced by temperature changes, wetting and drying or pressure release. Volumetric expansion of material within rock pores and cracks predominantly involves salt and ice crystals. Although mechanical weathering processes may cause rock disintegration by themselves, they often operate in association with CHEMICAL WEATHERING and biological weathering processes.

Expansion of rocks and minerals

A rock unit beneath the land surface is subjected to high compressive stress from the rock and sediment layers above. Once surface erosion removes the overlying rock layers, the rock unit will tend to expand in the direction of stress (or pressure) release. This may result in sheet jointing parallel to the unloading surface. As the rock surface expands in response to the PRESSURE RELEASE, the tensional strength of the rock may be exceeded by the tensile stress due to expansion, causing cracks to develop perpendicular to the rock surface. In this way, the rock unit may be broken into smaller slabs, which increases the surface area available for attack by other weathering processes. Pressure release cracking is most common in granites and metamorphic materials, and may also be seen on massive sandstones. As the surface sheets are lost, curved surfaces are formed; on a large scale (outcrop size) this process is termed exfoliation, on a smaller scale (boulder size) it is termed SPHEROIDAL WEATHERING. Exfoliation plays an important role in the creation of some landforms such as INSELBERGS, tors, arches and natural bridges.

Expansion of a rock surface creates tensile stresses and contraction of the surface creates compressive stresses. This may be induced by changes in the temperature of a rock surface. In a similar fashion to the stresses induced by pressure release, temperature-induced stresses decline in magnitude with increasing distance from the exposed rock surface. Stresses are restricted to the outer few centimetres due to the low conductivity of rock, which prevents inward transfer of heat (Hall and Hall 1991). Physical disruption of rock due to thermal shock occurs during forest fires, although the effectiveness of this process depends on rock composition (Ollier and Ash 1983). Whether or not receipt of insolation and diurnal temperature cycles can drive this process has long been debated. Daily temperatures in hot desert environments may reach 50 °C, and rock surface temperatures can reach 80 °C, with rapid cooling at night. Early anecdotes about rocks cracking in the desert were dismissed as hearsay when early experimental studies suggested the process was not viable. More recent research has suggested that thermal expansion, or insolation weathering, may indeed cause rock disintegration, though its effectiveness may largely be dependent on sufficient moisture within the rock. Igneous rocks, which contain many types of minerals with differing coefficients of thermal expansion, may experience stresses as a result of differential thermal response of minerals to heating and cooling cycles.

Rock minerals may expand when water is introduced into their structure; certain clay minerals typically behave in this way. Clays such as smectites and montmorillonites have the capacity to absorb water into the mineral during periods of wetting, which causes the mineral to swell. Bentonite (Na-montmorillonite), for example, may increase in volume by up to 1,500 per cent due to hydration and swelling. The species of clay mineral determines the degree to which it will expand and contract on wetting and drying (Yatsu 1988).

Expansion of material in voids

Exposed rock surfaces experience cycles of WETTING AND DRYING WEATHERING related to rainfall events and periods of evaporation. Simple wetting and drying of some rocks may cause their breakdown. When water enters a crack, or void, on a rock surface, it will become adsorbed by minerals lining the crack which may show unsatisfied electrostatic bonds. Further ingress of water may induce a swelling pressure within the void. Evaporation will remove all water molecules except those strongly bonded to the mineral surfaces; the sides of the crack, or void, may be pulled together by the attractive forces between water molecules on opposing faces. In this way, cycles of wetting and drying may induce expansion and contraction, which can split susceptible rocks such as shale, schist and even sedimentary rocks. Rocks that may be affected by wetting and drying usually have high clay mineral content, structural weaknesses, high permeability and low tensile strength. Wetting and drying may increase the size and/or number of pores and microcracks in rock, which has important implications for both frost (see FROST AND FROST WEATHERING) and SALT WEATHERING; an increase in water absorption capacity and a reduction in rock strength will accelerate the action of these processes.

Frost weathering involves the breakdown of rock as a result of the stresses induced by the freezing of water. Water experiences a 9 per cent increase in volume upon freezing and, in a closed system, may create pressures (theoretically 250 MPa) that exceed the tensile strength of rock (typically 25 MPa). A closed system may be produced if water in a rock freezes rapidly form the surface downwards, allowing ice to seal water in surface cracks and pores in the rock. Experimental work has shown that numerous mechanisms are involved in frost weathering, not only volumetric expansion of water; the most significant include adsorptive suction, as pore water moves toward the freezing front, and expansion (0.6 per cent from $+4\,°C$ to $-10\,°C$) of absorbed water. Many experimental studies have shown the importance of rapid freezing rates (at least $0.1\,°C$ per minute), low minimum temperatures ($<-5\,°C$) and high rock moisture content in determining the efficiency of frost weathering in causing shattering of rock samples (McGreevy and Whalley 1985). Moisture content is particularly important as the presence of air in pore spaces in unsaturated rock allows ice to expand into empty pores and voids and prevents crack growth. Rock properties exert an important control on the efficiency of frost weathering, as texture and structure determine both water absorption capacity and strength. Igneous and metamorphic rocks tend to be most resistant to frost shattering, while shales, sandstones and porous chalk tend to be least resistant. Frost weathering is likely to be most effective in alpine and cold temperate environments where freeze–thaw cycles occur frequently and abundant moisture is available, rather than cold deserts and polar areas. Angular rock fragments produced by frost shattering are termed felsenmeer.

Salts, the chemical compounds formed from reactions between acids and bases, are very important in causing rock breakdown. The effects of salt attack can best be seen in arid, coastal and urban environments, where salts are available and rocks routinely experience desiccating conditions that allow the salts to crystallize. The most common salts found in rocks are halite (NaCl), gypsum ($Ca_2SO_4 \cdot 2H_2O$), sodium sulphate (Na_2SO_4), magnesium sulphate ($MgSO_4$), sodium carbonate (Na_2CO_3) and sodium nitrate ($NaNO_3$) and their hydrated forms. The stresses causing rock breakdown are produced by three mechanisms: crystallization of salts in rock cracks and voids, expansion of salt crystals on hydration and thermal expansion of crystals (Cooke and Smalley 1968).

The most potent cause of salt weathering is salt crystal growth (Goudie and Viles 1997). Crystal growth occurs as a result of a saline solution becoming saturated as evaporation occurs and/or temperature changes, or by mixing of salts in solution, termed the 'common ion effect' (Goudie 1989). The role of salt crystallization in causing breakdown depends on the pore-size distribution and the pore connectivity of the material, as well as its overall strength. A second cause of stress arises from the capacity of certain common salts to take significant quantities of water into their structure. This hydration causes volumetric expansion of the salt and may exert pressure on crack and pore walls. Sodium sulphate, for example, will expand by 313 per cent on hydration. Hydration expansion may occur in response to changes in relative humidity, which, as this is closely related to temperature, may be diurnal. The extent to which salts expand when they are heated depends on their thermal characteristics

and the temperature ranges to which they are subjected. Most commonly occurring salts have coefficients of expansion higher than those of rocks, yet simulation studies have not demonstrated differential thermal expansion to be a very effective weathering mechanism in isolation from other effects of salts. One reason may be that rocks only experience thermal cycling in a shallow surface layer, so subsurface salts probably do not experience significant diurnal temperature changes.

Evidence exists for the severe damage caused by salts to many rock types in a range of environments. Weathering forms produced will vary with lithology, though landforms produced by weathering are usually small or minor forms. Salt weathering processes may produce CAVERNOUS WEATHERING, flaking, scaling and GRANULAR DISINTEGRATION of the surfaces of most rock types; porous sedimentary rocks are particularly susceptible. Decay to stone used in buildings and monuments is commonly caused by salt attack, induced by salt-rich environmental conditions (Cooke *et al.* 1993) and by polluted conditions in urban atmospheres (Cooke and Doornkamp 1990).

Other material which can expand in rock voids, causing internal stresses and eventual breakup, include plant material. Growth of plant roots or lichen thallus, for example, in cracks in rock may have a biophysical effect in creating growth stresses. The likely tensile stresses created are, however, smaller (3 MPa) than the tensile strength of most rocks. The impact of ORGANIC WEATHERING is the result of a complex suite of biochemical and biophysical processes, and cannot be explained by mechanical weathering processes alone.

References

Cooke, R.U., Brunsden, D., Doornkamp, J.C. and Jones, D.K.C. (1993) *Urban Geomorphology in Drylands*, Oxford: Oxford University Press.
Cooke, R.U. and Doornkamp, J.C. (1990) *Geomorphology in Environmental Management*, 2nd edition, Oxford: Oxford University Press.
Cooke, R.U. and Smalley, I.J. (1968) Salt weathering in deserts, *Nature* 220, 1,226–1,227.
Goudie, A.S. (1989) Weathering processes, in D.S.G. Thomas (ed.) *Arid Zone Geomorphology*, 11–24, London: Belhaven Press.
Goudie, A.S. and Viles, H. (1997) *Salt Weathering Hazards*, Chichester: Wiley.
Hall, K. and Hall, A. (1991) Thermal gradients and rock weathering at low temperatures: some simulation data, *Permafrost and Periglacial Processes* 2, 103–112.
McGreevy, J.P. and Whalley, W.B. (1985) Rock moisture content and frost weathering under natural and experimental conditions; a comparative discussion, *Arctic and Alpine Research* 17, 337–346.
Ollier, C.D. and Ash, J.E. (1983) Fire and rock breakdown, *Zeitschrift für Geomorphologie* 27, 363–374.
Yatsu, E. (1988) *The Nature of Weathering: An Introduction*, Tokyo: Sozosha.

Further reading

Amoroso, G.G. and Fassina, V. (1983) *Stone Decay and Conservation*, Materials Science Monographs 11, New York: Elsevier.
Bland, W. and Rolls, D. (1998) *Weathering: An Introduction to the Scientific Principles*, London: Arnold.
Ollier, C.D. (1984) *Weathering*, 2nd edition, London: Longman.
Whalley, W.B. and McGreevy, J.P. (1985) Weathering, *Progress in Physical Geography* 9, 559–581.
Whalley, W.B. and McGreevy, J.P. (1987) Weathering, *Progress in Physical Geography* 11, 357–369.
Whalley, W.B. and McGreevy, J.P. (1988) Weathering, *Progress in Physical Geography* 12, 130–143.

SEE ALSO: deep weathering; frost and frost weathering; honeycomb weathering; hydration; organic weathering; pressure release; salt weathering; wetting and drying weathering

ALICE TURKINGTON

MECHANICS OF GEOLOGICAL MATERIALS

Newton's laws of motion provide the basis for the science of classical mechanics, and the mechanics of geological materials involves mostly a special branch of classical mechanics called continuum mechanics. Continuum mechanics utilizes the same physical laws that govern motion of discrete bodies such as planets and billiard balls, but continuum mechanics addresses the internal deformation of bodies that cannot be idealized as discrete points.

Familiar geomorphological examples of deformable bodies that can be analyzed using continuum mechanics include water flowing in streams and soil moving down slopes. In treating such bodies as continua, a basic assumption is that the elemental constituents of the bodies (e.g. molecules of water or grains of soil) are minuscule relative to the scale of observable phenomena of interest. Observable motion of streams or slope debris typically involves momentum exchange amongst billions of water molecules or soil grains, and it is consequently impractical to analyse the movements of individual molecules or grains to predict the behaviour of the aggregate. A logical alternative is to analyse the aggregated behaviour of a mass of grains or molecules by treating them as a single deformable body, or 'continuum'.

Numerous textbooks provide excellent introductions to continuum mechanics. Examples range from mathematically rigorous treatises (e.g. Malvern 1997) to introductory books aimed at the interests of Earth scientists (e.g. Middleton and Wilcock 1994). An oft-overlooked but very informative book is the small classic by Jaeger (1971).

Continuum conservation laws

The central principles in geological continuum mechanics are conservation of mass, momentum and energy. Although both momentum and energy conservation apply in every geomorphological setting, momentum conservation is generally the more useful principle, because momentum is a vector quantity that includes information about the direction of motion, whereas energy is a scalar. However, if thermal effects or phase changes are important (as in melting or formation of ice, for example) energy conservation must be considered explicitly, in addition to conservation of momentum and mass. The discussion below assumes that thermal effects are negligible, and emphasizes purely mechanical behaviour that arises solely from conservation of momentum and mass.

The basic equations of mass and momentum conservation describe behaviour in four dimensions (space + time), and they apply to any continuous body, regardless of its composition or state (solid, liquid or gas). The equations can be written in mathematical vector notation (with a brief English translation beneath) as:

mass conservation: $$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \vec{v} = 0$$

local rate of mass increase + rate of mass efflux due to deformation = 0

momentum conservation: $$\rho \frac{\partial \vec{v}}{\partial t} + \rho \vec{v} \nabla \cdot \vec{v} = \rho \vec{g} + \nabla \cdot T$$

local rate of momentum increase + rate of momentum efflux due to deformation = force imposed by gravity + internal reaction force (stress).

In these equations the dependent variables are ρ, the local mass density within the continuous substance, and $\vec{v}$, the local vector velocity, which can vary as a function of position and time, t. In many geomorphological phenomena, the only imposed driving force is the so-called 'body force' $\rho\vec{g}$ due to gravitational acceleration, $\vec{g}$. Additional imposed forces can be included if necessary.

The final quantity in the momentum-conservation equation is the stress, T, which represents the reaction forces (per unit area) that develop within a deformable body as a consequence of the interaction between the driving force and local accelerations. Unlike a rigid body, in which the action-reaction involving gravitational driving force and acceleration can be represented by a familiar form of Newton's second law, $\rho(d\vec{v}/dt) = \rho\vec{g}$, a deformable, continuous body can react to external forcing by generating internal stress. Thus, the concept of stress is a key one in continuum mechanics, and it should be mastered by all students of physical geomorphology. In general stress is a second-order tensor quantity (commonly represented mathematically by a 3×3 matrix), and the stress tensor is symmetric (which eliminates the need for a separate angular momentum equation in addition to the linear momentum equation above). Textbooks that explain stress and tensors in the context of geological sciences include an excellent one by Means (1976).

An important general feature of stress is known as 'static indeterminacy'. This condition dictates that even if a continuous substance is motionless and the momentum equation above reduces to $\rho\vec{g} + \nabla \cdot T = 0$, the stresses cannot be calculated without specifying a 'constitutive equation' that summarizes the mechanism of stress generation. The only mechanical analyses in which static indeterminacy and the need for constitutive equations can be circumvented are analyses in which stress is assumed to vary in one direction only. These 'one-dimensional' analyses can be useful for building insight but seldom provide accurate models of multidimensional geomorphic phenomena.

Constitutive equations and rheology

Two main branches of continuum mechanics, solid mechanics and fluid mechanics, have developed from observations of the fundamentally different mechanisms by which solids and fluids (liquids and gases) generate stress. In solid mechanics a pivotal observation is that, for sufficiently small deformations, stress is simply proportional to the magnitude of deformation (or, more precisely, proportional to the magnitude of strain, which may differ subtly from deformation). First quantified by Robert Hooke (1635–1703), this observation led to the theory of linear (or 'Hookean') elasticity. A similarly pivotal postulate, first made by Isaac Newton (1642–1727), was that fluids deform such that stress is simply proportional to the *rate* of deformation. Subsequently confirmed by experiments, this postulate led to the linear (or 'Newtonian') theory of viscous fluid flow.

Equations that relate stress to deformation are known as constitutive equations, because they express the influence of a body's constitution (or rheology) on the internal reaction forces that generate stress. In effect, constitutive equations (i.e. rheological models) serve as surrogates for momentum exchange that occurs at scales too small to be resolvable in a particular setting or observation (for example, momentum exchange by colliding water molecules in a stream observed by eye). The constitutive equations for linearly elastic solids and linearly viscous fluids may be written in simple forms as

$$T = \epsilon D \text{ (linearly elastic behaviour)}$$

$$T = \eta \dot{D} \text{ (linearly viscous behaviour)}$$

where ϵ represents elastic moduli, η represents dynamic viscosity, D represents deformation, and $\dot{D}$ represents deformation rate. Whole treatises expound the detailed meaning of these equations and the detailed definition of D and $\dot{D}$ (which are tensor quantities, like stress). Here it suffices to emphasize that these well-known constitutive equations express straightforward connections between stress and a measurable macroscopic quantity such as D or $\dot{D}$.

Stresses in Earth-surface fluids such as water and air can be represented with fair accuracy by a simple constitutive equation describing linearly viscous behaviour, and stresses in a solid rock can be represented with similarly good accuracy by linearly elastic behaviour (provided the rock does not fracture). However, many materials encountered in geomorphology are not so simple. Rocks, soils and sediments that undergo large, irreversible deformation (as might occur in a landslide, for example) exhibit neither linearly viscous nor linearly elastic behaviour. Diverse constitutive equations have been proposed to represent such behaviour, but the discussion here introduces only the most significant equation, the Coulomb model.

Experiments demonstrate that stresses in soils, sediments and fragmented rocks undergoing large deformations adjust plastically to maintain limiting values, independent of the deformation magnitude or rate. These limiting 'plastic yield' stresses depend principally on friction due to rubbing and interlocking of adjacent grains, and to a lesser degree on cohesive bonding between grains. On planes of shearing the limiting shear stress τ is described by

$$\tau = \sigma \tan \phi + c \text{ (Coulomb plastic behaviour)}$$

where σ is the normal stress on the shear plane, φ is the angle of internal friction, and c is the cohesion. This equation, first posited by Charles Augustin Coulomb (1736–1806), has withstood repeated testing but does not provide the same straightforward interpretation as the linearly elastic and viscous equations noted above. Nonetheless, the Coulomb model has proved very useful for analysing phenomena such as LANDSLIDES, DEBRIS FLOWS and incipient motion of BEDLOAD particles in streams.

Initial and boundary conditions

In addition to conservation laws and constitutive equations, continuum mechanics requires specification of initial conditions that isolate a phenomenon in time as well as boundary conditions that isolate it in space. (For example, to mechanically analyse the behaviour of a flood, one must specify the initial channel geometry and the distributions of water-surface elevations and velocities prior to the flood onset.) Appropriate specification of these 'auxiliary conditions' can be the crux of successful mechanical modelling, because geomorphological phenomena seldom occur in isolation from the surrounding environment. Nonetheless, mechanical analyses of such 'open' geomorphological systems can provide key insights if auxiliary conditions are specified with care and precision (Iverson 2003).

References

Iverson, R.M. (2003) How should mathematical models of geomorphic processes be judged? in P.R. Wilcock and R.M. Iverson (eds) *Prediction in Geomorphology*, Washington, DC: American Geophysical Union.

Jaeger, J.C. (1971) *Elasticity, Fracture and Flow with Engineering and Geological Applications*, London: Chapman and Hall.
Malvern, L.E. (1997) *Introduction to the Mechanics of a Continuous Medium*, 2nd edition, Englewood Cliffs: Prentice-Hall.
Means, W.D. (1976) *Stress and Strain (Basic Concepts of Continuum Mechanics for Geologists)*, New York: Springer-Verlag.
Middleton, G.V. and Wilcock, P.R. (1994) *Mechanics in the Earth and Environmental Sciences*, Cambridge: Cambridge University Press.

RICHARD M. IVERSON

MEGAFAN

Large fluvial depositional features that have been defined thus (Horton and DeCelles 2001: 44):

> Fluvial megafans constitute volumetrically significant depositional elements of sedimentary basins adjacent to mountain belts. A fluvial megafan is a large ($10^3-10^5\,km^2$), fan-shaped (in plan view) mass of clastic sediment deposited by a laterally mobile river system that emanates from the outlet point of a large mountainous drainage network. Modern fluvial megafans have been recognized in nonmarine foreland basin systems at the outlets of major rivers that drain fold-thrust belts, particularly in the Himalayas and northern Andes. Although fluvial megafans are similar to sediment gravity flow-dominated and stream-dominated alluvial fans in terms of their piedmont setting, planform geometry and sedimentation related to expansion of flow downslope of a drainage outlet, fluvial megafans are distinguished by their greater size (alluvial fans rarely exceed $250\,km^2$), lower slope, presence of floodplain areas and absence of sediment gravity flows. The term 'terminal fan' is commonly used for a large, distributary fluvial system in which surface water infiltrates and evaporates before it can flow out of the system. A fluvial megafan therefore may be considered 'terminal' in cases where fluvial channels run dry before reaching bodies of water downstream. Although fluvial megafans are clearly related to the emergence of a large mountain river onto a low-relief alluvial plain, their stratigraphic evolution in nonmarine foreland basins may also be critically dependent on variables such as sediment flux, water discharge, drainage catchment size, catchment lithology and subsidence rate, factors that are ultimately controlled by tectonic, climatic and geomorphic processes.

Alternative names include megacone, inland delta, wet alluvial fan and braided stream fan. Some especially impressive megafans occur on the north side of the Ganga plain in India. They are fed from the Himalayas and may have formed in the Late Pleistocene when coarser grained sediment and high sediment and water discharges were available (Shukla *et al.* 2001).

Not all megafans occur in such dramatic settings as those of the Andean and Himalayan forelands. For example, the Okavango Fan of northern Botswana has accumulated in a graben and has been deposited by a low sinuosity/meandering river rather than a braided stream system (Stannistreet and McCarthy 1993).

References

Horton, B.K. and DeCelles, P.G. (2001) Modern and ancient fluvial megafans I. The foreland basin system of the Central Andes, southern Bolivia: implications for drainage network evolution in fold-thrust belts, *Basin Research* 13, 43–63.
Shukla, U.K., Singh, I.B., Sharma, M. and Sharma, S. (2001) A model of alluvial megafan sedimentation: Ganga Megafan, *Sedimentary Geology* 144, 243–262.
Stannistreet, I.G. and McCarthy, T.S. (1993) The Okavango Fan and the classification of subaerial fan systems, *Sedimentary Geology* 85, 115–133.

A.S. GOUDIE

MEGAGEOMORPHOLOGY

The term 'mega-geomorphology' was introduced in 1981 to designate a London conference of the British Geomorphological Research Group (Gardner and Scoging 1983). The original intent was for the term to apply to geomorphology on the scale of plate tectonics, biological evolution and macro-climatic change. Thus, it was to be concerned with entire landscapes, through histories of millions of years, in the context of continental or macro-regional evolution.

In 1985 another conference, 'Global Megageomorphology', was organized in Oracle, Arizona, and this meeting further refined geomorphological study at the largest spatial and temporal scales (Baker and Head 1985). Particularly relevant issues for the Oracle conference were the use of orbital remote sensing procedures to produce global mapping and analyses, studies of

continental-scale denudation, the relationship of geomorphology to regional tectonics, global environmental change, and the geomorphology of other rocky planets besides Earth. The last issue is particularly important in both philosophical and historical perspectives. The history of geomorphological study of Earth began with scientists observing their immediate surroundings, and then generalizing to explanations that placed those observations in a larger context. In contrast, the study of other rocky planetary surfaces always begins at the largest spatial scales, through the remote sensing instrumentation of flyby planetary missions. The geomorphological understanding of Mars and Venus, the most Earthlike of the known planets, began at the megascale, while understanding of Earth's surface began at small scales and only evolved to the megascale after a long history of observations at the scales that were most accessible to human observers. A result of this history is the set of theories that currently constrains geomorphological inquiry to somewhat limiting viewpoints (Baker and Twidale 1991; Baker 1993). Megageomorphology affords an opportunity to break the constraints and develop new sets of theories.

The combined issues of scale and viewpoint are highlighted by a survey in the United Kingdom that found for the early 1980s that 75 per cent of geomorphological research concerned small-scale, modern process studies and another 15 per cent concerned Quaternary studies (Gardner and Scoging 1983). This emphasis on small scales of time and space is indicative of a reductionistic epistemological perspective in which one presumes that small-scale studies will integrate over time to generate a theory of the whole. Small-scale, modern process studies also tend to minimize the geomorphological role of rare processes of extreme magnitude because these are both remote from possible direct observation and destructive of most attempts to measure (Baker 1988). In contrast, as shown in Figure 107, the processes responsible for landforms and landscapes operate over a broad range of spatial and temporal scales. Moreover, the responses to these processes add further extensions of time and space to the zone of relevant natural operation for geomorphological change. One must conclude, 'Contemporary process studies are of little worth in evaluating landscape evolution' (Church 1980).

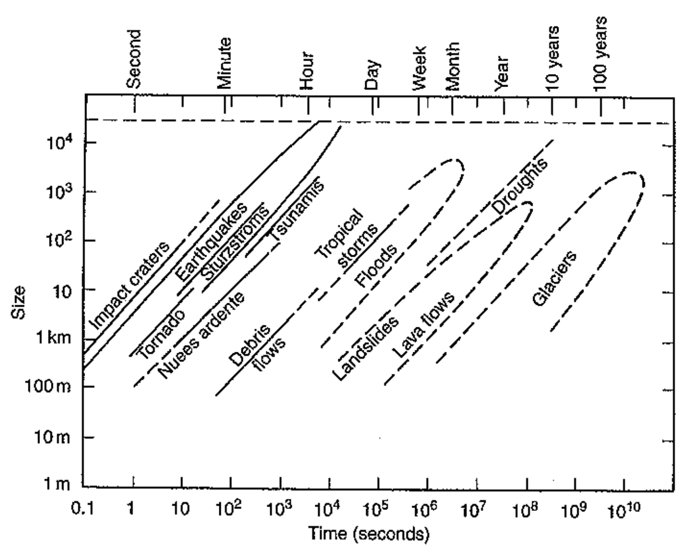

Figure 107 Scale relationships for various geomorphological processes, based on Carey (1962)

Many geomorphological systems are nonlinear, with thresholds that involve negligible responses at the process scales most easily measured in the field, while the relevant processes are both rare and of extreme magnitude (Baker 1988). The debates over the origin of the Channeled Scabland in eastern Washington in the 1920s illustrate this issue because a very large-scale process of immense magnitude was involved (Baker 1978). It is only because the process of megaflooding was recognized and understood for the Channeled Scabland that subsequent work has been able to show that cataclysmic megafloods played the dominant role in landscape development for other parts of Earth, and, somewhat surprisingly, for parts of Mars as well (Baker 2002).

Modern megageomorphology makes extensive use of global observations from spacecraft that employ a variety of imaging and remote-sensing instrumentation, including multispectral imaging, radiometers and radars. Image processing of digitally formated data has revolutionized the ability to study landscapes at the largest spatial scales. The theme of global-scale remote sensing is developed in the book *Geomorphology from Space* (Short and Blair 1986). A full scheme of geomorphology at large spatial scales is readily achieved in regard to the new technologies (Baker 1986).

Other technological advances that are stimulating megageomorphology include the quantitative geochronology of geomorphic surfaces and deposits; the widespread availability of digital topography; and the mathematical modelling of landscape evolution. These elements are creatively employed in the rapidly developing subfield of tectonic geomorphology (Burbank and Anderson 2001). In essence, the tectonic geomorphologist performs thought experiments with the computer, then tests those notions against the topographic response and the ages of elements in the landscape, such as glacial moraines, stream terraces and denudation surfaces. This can be done for entire regions that comprise the major tectonic elements of the planet.

The two major factors for geomorphological evolution at very large scales are tectonic factors arising from forces inside Earth (with the occasional imposition of extraterrestrial impacts) and the denudational factors arising from Earth's atmosphere, as largely summarized in its climate. Both these concerns are critical to any modern sense of global geomorphology (Summerfield 1991). The understanding of Earth tectonics was revolutionized in the 1960s and 1970s by the plate-tectonic theory. During the 1980s and 1990s climate became the central issue for international initiatives to understand global environmental change and the operation of Earth as a system. With the theoretical underpinnings of these two major elements, megageomorphology is now poised for major development as a science.

It is particularly interesting that a form of megageomorphology developed after 1945 in the former Soviet Union. Because of its practical application to mineral and petroleum exploration, the details of this science were subject to state secrecy, and its community of practitioners became somewhat isolated from the greater world scientific community. The focus of this science was the study of morphostructures, which are linear and circular elements of continental landscapes controlled by tectonics and denudation. The morphostructures exhibit hierarchical relationships with one another, and they evolve over geological timescales (Baker *et al.* 1993). Morphostructural analyis deciphers the complex interaction of long-term endogenetic processes with surface relief. It proved especially useful in identifying various concentric, circular or oval, and linear hidden dislocations of basement rocks that on Earth are commonly obscured by sedimentary or volcanic cover, deformations and intrusions. Though not tied to the modern plate-tectonic ideas about Earth's large-scale evolution, morphostructures are surprisingly similar to various quasi-circular shaped upland regions discovered on Venus by the planetary missions of the 1980s and 1990s. These also exist in hierarchical arrangement and may be related to mantle plume tectonics (Finn *et al.* 1994). Merely thinking at the megascale led to some surprising results.

Sharp (1980: 231) observes, 'One of the lessons from space is to "think big"'. This theme, combined with new analytical tools for geomorphological study at very large scales affords an opportunity for discovery and scientific excitement. Unifying models of global tectonics and climate dynamics afford the scientific framework for large-scale studies. Orbital remotes sensing, digital topographic data and geochemical tools for dating Earth history all permit the quantification of geomorphological parameters in greatly expanded temporal and spatial domains. Moreover, the science is immensely stimulated by the discovery of entire new landscapes on the surfaces of other rocky planetary bodies.

References

Baker, V.R. (1978) The Spokane Flood controversy and the Martian outflow channels, *Science* 202, 1,249–1,256.
——(1986) Introduction: regional landforms analysis, in N.M. Short and R.W. Blair, Jr (eds) *Geomorphology from Space: A Global Overview of Regional Landforms*, 1–26, Washington, NASA Special Publication 486.
——(1988) Cataclysmic processes in geomorphological systems, *Zeitschrift für Geomorphologie Supplementband* 67, 25–32.
——(1993) Extraterrestrial geomorphology: science and philosophy of Earthlike planetary landscapes, *Geomorphology* 7, 9–35.
——(2002) High-energy megafloods: planetary occurrences and sedimentary dynamics, in I.P. Martini, V.R. Baker and G. Garzon (eds) *Flood and Megaflood Processes and Deposits: Recent and Ancient Examples*, 3–15, International Association of Sedimentologists Special Publication 32.
Baker, V.R. and Head, J.W., III (1985) Global megageomorphology, in R.S. Hayden (ed.) *Global Geomorphology*, 113–120, Washington, NASA Conference Publication 2,312.
Baker, V.R. and Twidale, C.R. (1991) The reenchantment of geomorphology, *Geomorphology* 4, 73–100.
Baker, V.R., Finn, V.I. and Komatsu, G. (1993) Morphostructural megageomorphology, *Israel Journal of Earth Sciences* 41, 65–73.
Burbank, D.W. and Anderson, R.S. (2001) *Tectonic Geomorphology*, Oxford: Blackwell.
Carey, S.W. (1962) Scale of geotectonic phenomena, *Journal of the Geological Society of India* 3, 97–105.
Church, M.A. (1980) Records of recent geomorphological events, in D. Cullingford, D.A. Davidson and J. Lewin (eds) *Timescales in Geomorphology*, 13–29, New York: Wiley.
Finn, V.S., Baker, V.R., Dolginov, A.Z., Gabitov, I. and Dyachenko, A. (1994) Large-scale spatial patterns in topography at Alpha Regio, Venus, *Geophysical Research Letters* 22, 1,901–1,904.
Gardner, R. and Scoging, H. (eds) (1983) *Mega-Geomorphology*, Oxford: Clarendon Press.
Sharp, R.P. (1980) Geomorphological processes on terrestrial planetary surfaces, *Annual Reviews of Earth and Planetary Sciences* 8, 231–261.
Short, N.M. and Blair, R.W., Jr (eds) (1986) *Geomorphology from Space: A Global Overview of Regional Landforms*, Washington, NASA Special Publication 486.
Summerfield, M.A. (1991) *Global Geomorphology*, Harlow, Essex: Longman.

SEE ALSO: extraterrestrial geomorphology; global geomorphology; tectonic geomorphology

VICTOR R. BAKER

MEKGACHA

A Setswana term for the DRY VALLEY systems which traverse the flat, sandy terrain of the Kalahari region of southern Africa. These broad, shallow, drainage features contain CALCRETE and SILCRETE in their floors and flanks and are also referred to as *laagte*, *omuramba* or *dum* in other regional languages. The origins of *mekgacha* (singular *mokgacha*) have been ascribed to episodes of permanent or ephemeral fluvial activity during wetter periods of the Quaternary, and, at longer timescales, to RIVER CAPTURE, groundwater sapping or DEEP WEATHERING focused along geological lineaments (Nash *et al.* 1994). Two groups of valleys can be identified (Shaw *et al.* 1992). The first are the exoreic (externally directed) Auob, Nossop, Kuruman and Molopo systems which drain the southern Kalahari and connect to the Orange River. The second are the endoreic (internally directed) systems which focus upon the Okavango Delta and Makgadikgadi Depression. Surface runoff is comparatively rare within the Kalahari and, as a result, flow within *mekgacha* is unusual. However, most exoreic systems are spring-fed and contain water in their headwater sections, with more extensive flooding following prolonged rainfall. In contrast, the endoreic systems are effectively fossil networks and only contain water in seasonal pools. Floods may occur in such networks after exceptional rainfall, but only two have been documented in the historical period (1851 and 1969), both in the Letlhakane valley (Nash and Endfield 2002).

References

Nash, D.J. and Endfield, G.H. (2002) Historical flows in the dry valleys of the Kalahari identified from missionary correspondence, *South African Journal of Science* 98, 244–248.
Nash, D.J., Thomas, D.S.G. and Shaw, P.A. (1994) Timescales, environmental change and valley development, in A.C. Millington and K. Pye (eds) *Environmental Change in Drylands*, 25–41, Chichester: Wiley.
Shaw, P.A., Thomas, D.S.G. and Nash, D.J. (1992) Late Quaternary fluvial activity in the dry valleys (*mekgacha*) of the Middle and Southern Kalahari, southern Africa, *Journal of Quaternary Science* 7, 273–281.

DAVID J. NASH

MELTWATER AND MELTWATER CHANNEL

Meltwater can be produced from the melting of snow (nival meltwater) or GLACIERs (glacial meltwater). The nival and glacial meltwater regimes represent two ends of a spectrum, as in many instances there may be contributions of both within the same catchment. The amount of meltwater produced is determined by the energy balance, a key component of which is solar radiation, thus most melt will occur at the ice or snow surface. Melting can also occur at the base of a glacier due to either geothermal heat or the effects of high pressure. However, melting will predominantly be related to air temperature, thus the temporal pattern of meltwater generation is not constant and varies over daily, annual and longer timescales. In nival regimes highest flows tend to occur in spring (the start of the melt season) as temperatures begin to rise above freezing. Meltwater discharge will usually exhibit a diurnal pattern, with a peak in the afternoon and a low around dawn, related to the cycle of daily temperature change. Over time, as less snow remains to be melted, diurnal fluctuations become more muted and flow will eventually cease altogether. Glacial meltwater production also exhibits a similar type of pattern, except that the melt season is longer and peak meltwater discharge is later. Glacial meltwater discharge displays a gradual rise through the melt season as the melting of snow on the ice surface is then followed by glacier melt. Peak flows in mid/late summer when the entire glacier, devoid of the insulating surface snow cover, contributes to melt. Outside the summer melt season flows can become very low or cease altogether. The specific meltwater regime is very closely tied with the environment, for example, glaciers in more polar environments will tend to have a shorter melt season and more muted diurnal discharge signature than those of mountain glaciers in temperate locations (Tranter *et al.* 1996). In addition to these more regular and predictable discharge regimes, meltwater can also be released catastrophically in OUTBURST FLOODs.

Once snow or ice has melted it can take several pathways with respect to its exit from an ice mass: as surface (supraglacial) channels, pipes or conduits within (englacial) the ice or as flow at the base (subglacial) of the ice (Shreve 1972). Supraglacial channels can be up to a few metres deep and because the ice is smooth water velocities are high (3–7 ms^{-1}). Supraglacial water will either flow off the ice surface or descend vertically into the ice via holes called MOULINs where the water connects with the pipes or conduits of the englacial system. Englacial pipes range in diameter from a few millimetres to a few metres, water can also flow through thin veins between ice crystals, but at a greatly reduced rate. Englacial water will often connect to the subglacial flow system at the base of the glacier. Where a cold-ice basal thermal regime predominates the supraglacial drainage system is often dominant, in a warm-ice situation a complex system involving all three types of drainage is often present. However, the drainage system is not fixed within the ice mass and will change and develop through the melt season as some channels or pipes open up and others close. In addition to this pattern of flow routes, meltwater can also be stored in supraglacial, englacial or subglacial lakes and ponds.

Subglacial meltwater has the biggest significance in terms of glacier dynamics and also in terms of providing evidence of meltwater flow in the landscape once the ice has receded. Subglacial meltwater significantly increases the rate of basal sliding and hence how fast the ice will move. However, if there is a highly permeable substrate at the base of the ice then very little meltwater flow may occur at the ice–bed interface, in which case basal sliding will not be greatly enhanced. A key factor about subglacial meltwater (and also englacial) is that due to the pressure of the weight of the overlying ice, the water can be flowing at much greater pressures than would be experienced under normal atmospheric pressure conditions at the surface. The direction of flow is determined by the ice surface slope and to a lesser extent the bed topography, such that uphill flow is possible (Shreve 1972). As the ice surface steepens, subglacial flow becomes less influenced by bed topography. Flow at the base will thus trend in the general direction of ice flow, often following valley floors and crossing divides at the lowest point.

Subglacial meltwater flow can occur in a distributed or discrete system. Distributed systems include sheet flow or linked cavity networks, discrete systems encompass the full range of channelized flows. Sheet flow is where meltwater exists as a thin continuous layer between the ice and the bed. Research has shown that this type of flow is relatively unstable and that flow is more likely to occur in channels (Shreve 1972). The linked cavity system is where basal hollows are linked by narrow short connections. Channelized systems can be cut down into the bed (Nye channels or N-type), cut up into the overlying ice (Röthlisberger channels or R-type) or display characteristics of both. The discrete (channelized) drainage systems transport meltwater more efficiently than the more circuitous pathways of the distributed system. The type of basal meltwater system thus has an important influence on water pressure and hence glacier motion. Meltwater can switch between these different flow systems and it has been suggested that this can trigger GLACIER surging.

Nye channels cut into bedrock or consolidated sediments leave the most distinctive imprint on the land surface once the ice has retreated. Due to the pressure conditions experienced by subglacial meltwater and the nature of meltwater supply to the subglacial system Nye channels often display very different characteristics to a normal subaerial fluvial system. For example, they can possess an undulating long profile, very steep gradients, lack any significant drainage basin and have an abrupt inception or termination (Glasser and Sambrook Smith 1999). The nature of the substrate beneath the ice will have an important influence on the overall channel morphology. For example, it is not uncommon for a channel to change its morphology over very short distances as it goes from being deep, narrow and incised in bedrock and then wide and shallow when passing over a deformable bed (Plate 79). Channels can range greatly in size, from tens of metres up to 100 kilometres long or from tens of metres to several kilometres wide. Once the channels reach the kilometre scale they are often referred to as TUNNEL VALLEYs rather than Nye channels. Channels can occur as either isolated features or part of a much larger channel network. The most spectacular form of isolated meltwater channel is what is referred to as a chute channel; these often occur on the flanks of a bedrock slope and are very steep and incised. They are thought to form as rapidly descending water within an ice mass reaches the glacier bed and cuts down into it. In contrast to these isolated channels, networks of large tunnel valleys can extend over very large areas and have been reported from many parts of North America and Europe that were covered by ICE SHEETs. The origin of such large features is thought to be due to either the catastrophic release of water stored in large subglacial lakes or that they were cut by normal meltwater flows over extended time periods. By mapping networks of exposed meltwater channels and coupling this with the theory of meltwater flow under pressure it has been demonstrated that reconstructions of the likely extent and dynamics of an ice mass is possible (Sugden *et al.* 1991).

Plate 79 A meltwater channel (flow away from camera) incised ~7 m into bedrock, Cheshire, England. Downstream the channel becomes wider and shallower as it passes over unconsolidated sediment before disappearing completely 500 m from where this photograph was taken. Photograph taken by N.F. Glasser

As well as being erosive, meltwater can also be a significant agent of deposition. For example, an ESKER is a long narrow ridge of sediment often deposited in subglacial channels as a result of high rates of meltwater flow. When subglacial channels completely fill with meltwater they can behave like pipes, allowing water under pressure to move, and subsequently deposit, large volumes of sediment in a single event.

References

Glasser, N.F. and Sambrook Smith, G.H. (1999) Glacial meltwater erosion of the Mid-Cheshire Ridge: implications for ice dynamics during the Late Devensian glaciation of northwest England, *Journal of Quaternary Science* 14, 703–710.

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

Sugden, D.E., Denton, D.H. and Marchant, D.R. (1991) Subglacial meltwater channel system and ice sheet overriding of the Asgard range, Antarctica, *Geografiska Annaler* 73A, 109–121.

Tranter, M., Brown, G.H., Hodson, A.J. and Gurnell, A.M. (1996) Hydrochemistry as an indicator of subglacial drainage system structure: a comparison of alpine and sub-polar environments, *Hydrological Processes* 10, 541–556.

Further reading

Menzies, J. (1995) Hydrology of glaciers, in J. Menzies (ed.) *Modern Glacial Environments*, 197–239, Oxford: Butterworth-Heinemann.

Menzies, J. and Shilts, W.W. (1996) Subglacial environments, in J. Menzies (ed.) *Past Glacial Environments*, 15–136, Oxford: Butterworth-Heinemann.

SEE ALSO: glacifluvial; subglacial geomorphology

GREGORY H. SAMBROOK SMITH

MESA

Mesa (the word is of Spanish origin) is a steep-sided and flat-topped hill or ridge rising above a flat plain, usually built of flat-lying soft sedimentary rocks capped by a more resistant layer, e.g. of shales overlain by sandstones. In volcanic terrains a former lava flow may act as a caprock; likewise, this role may be assumed by a blanket of DURICRUST. Mesas originate due to unequal scarp retreat, in the course of which parts of the plateau become isolated and remain standing in front of the retreating scarp. Lava and duricrust-capped mesas may form due to dissection of a plateau or through the mechanism of relief inversion (see INVERTED RELIEF), then the distribution of mesas indicates the position of a former valley floor or lava flow.

Mesas steadily decrease in size through slope retreat accomplished mainly by various kinds of mass wasting and gully erosion; however, due to the presence of a resistant cap they may be very durable landforms surviving long after the initial scarp retreated. With time, they are reduced to BUTTEs, but there are no agreed criteria as to when a mesa becomes a butte. Mesas are climate-independent landforms although those in desert areas have the most distinctive appearance.

SEE ALSO: caprock; sandstone geomorphology; structural landform

PIOTR MIGOŃ

METHOD OF SLICES

Slope stability is at present routinely analysed by the Limit Equilibrium approach, derived somewhat loosely from the Theory of Plasticity. It is based on the assumption that the pattern of stress in a failing slope can be determined from static equilibrium, without the need to consider stress redistribution due to elastic and inelastic straining. The stresses acting on the boundary of the sliding body ('rupture surface') can then be compared with available strengths, to evaluate equilibrium. The 'FACTOR OF SAFETY' is usually defined as a ratio by which available soil or rock strength may be reduced, without causing failure.

The Limit Equilibrium approach was originally applied to the sliding of rigid portions of a slope, along assumed rupture surfaces. Coulomb was the first to calculate the Factor of Safety of a block above a planar surface. Later, circular ('Swedish Circle') surfaces were analysed, as it was observed that slides in clay are often rotational.

In the 'Method of Slices' (Fellenius 1927), the sliding body, viewed in cross section, is divided into vertical slices. An approximate assumption is made that the vertical stress at the base of each slice is constant and equal to the weight of the material column above it. In the 'Ordinary' ('Fellenius') method, all stresses acting at the vertical boundaries between adjacent columns are neglected. The equilibrium problem then becomes statically determinate and could be solved by simple vector analysis. The Factor of Safety is equal to the ratio of the sum of available strengths on all column bases, to the sum of applied shear stresses. The results are often excessively conservative.

Bishop (1955) realized that, with a circular rupture surface, it is unlikely for the inter-slice *shear* stresses to be very high. He therefore assumed these to be zero and derived the normal and shear stresses at the base of each column from vertical force equilibrium. The Factor of Safety can then be evaluated from the moment equilibrium of the slice assembly, without the need to neglect the normal inter-slice forces. This method neglects horizontal force equilibrium, but nevertheless produces very accurate results when compared with more sophisticated approaches, for circular and some non-circular surfaces.

More detailed methods have been derived, taking into consideration inter-slice shear and all three equilibrium conditions. Recently, a much more sophisticated approach, based on numerical stress–strain analysis of the slope, has been developed, so called 'Stress Reduction Method' (Dawson *et al.* 1999). One of the advantages of this method is that it removes the need to predetermine the shape of the rupture surface. However, it is much more difficult to implement and lacks the long track record of practical experience, inherent in the Method of Slices. Thus, for the foreseeable future, the Method of Slices will continue to be an important tool for the analysis of slope stability. An extension to a three-dimensional 'Method of Columns' has now also been developed (e.g. Hungr *et al.* 1989).

References

Bishop, A.W. (1955) The use of the slip circle in the stability analysis of slopes, *Géotechnique* 5(1), 7–17.

Dawson, E.M., Roth, W.H. and Drescher, A. (1999) Slope stability analysis by strength reduction, *Géotechnique* 49, 835–840.

Fellenius, W. (1927) *Erdstatische Berechnungen mit Reibung und Kohasion*, Berlin: Ernst.

Hungr, O., Salgado, F.M. and Byrne, P.M. (1989) Evaluation of a three-dimensional method of slope stability analysis, *Canadian Geotechnical Journal* 27, 679–686.

OLDRICH HUNGR

MICRO-EROSION METER

Many theories of landform evolution over time can only be tested through a knowledge of erosion rates. Earlier ideas of erosion sequences were constructed through logic and deduction from available morphological and sedimentary evidence. However, the answer to the question: 'could the landform have actually evolved in the timescale envisaged?' must involve a knowledge of rates. Two sets of endeavours at different scales devolve from this question, first large-scale measurements, for example on drainage basins, and smaller scale measurements. The latter may be performed more precisely and over a short space of time but the issue then remains of how far the results may be influenced by conditions at the time of measurement and of how they may be scaled up over a longer time span and a larger scale. Thus many of the data derived have been useful in short-term experimentation but extrapolation over longer time spans, over which conditions may be different to those obtaining at the present, remains an issue (Trudgill *et al.* 1989, 2001).

It was realized by High and Hanna (1970) that a micrometer dial gauge, used in engineering, could provide measurements of up to 0.0001 mm or even greater precision. A measurement is made of the height of a rock surface relative to some fixed datum. The instrument consists of a micrometer dial gauge and attached micrometer probe which is mounted onto a tripod framework. The tripod gives the instrument stability and it rests on three reference studs drilled securely into the rock. The measurement of the surface height of the rock relative to the studs can be made at repeated intervals, yielding results of surface lowering in mm a^{-1}. Initial meters had probes which were fixed and the tripod could be rotated to three positions, yielding three measurement points. Later meters (Trudgill *et al.* 1981; Trudgill 1983) used a traversing mechanism where not only was there a tripod base plate which could be rotated, additionally the dial gauge could itself be placed in several reference positions enabling a much larger number of points to be measured.

In terms of methodological limitations, Spate *et al.* (1985) suggest that erosion of the rock by the tip of the probe could lead to a fictitious annual loss of around 0.019 mm a^{-1}. They reported that when repeated measurements were made successively then surface lowering was recorded, the more so in softer limestones. For replicate readings at one site, on a harder (Buchan) limestone their data showed differences of 0.0001–0.0052 mm and for a softer (Gambier) limestone 0.0090–0.0284 mm. This may be initial surface compaction rather than actual erosion. Their data suggest that up to 0.0126–0.0284 for softer limestones and 0.0016–0.0052 for harder limestones will be an artefact of probe impact for any one measurement rather than actual erosion.

References

High, C.J. and Hanna, K. (1970) A method for the direct measurement of erosion on rock surfaces, *British Geomorphological Research Group, Technical Bulletin* 5.

Spate, A.P., Jennings, J.N., Smith, D.I. and Greenaway, M.A. (1985) The micro-erosion meter: use and limitations, *Earth Surface Processes and Landforms* 10, 427–440.

Trudgill, S.T. (1983) *Weathering and Erosion*, 104–110, London: Butterworth.

Trudgill, S.T., High, C.J. and Hanna, K.K. (1981) Improvements to the micro-erosion meter, *British Geomorphological Research Group, Technical Bulletin* 29, 3–17.

Trudgill, S.T., Viles, H.A., Inkpen, R.J. and Cooke, R.U. (1989) Remeasurements of weathering

rates, St. Paul's Cathedral, London, *Earth Surface Processes and Landforms* 14, 175–196.
Trudgill, S.T., Viles, H.A., Inkpen, R., Moses, C., Goshing, W., Yates, T. *et al.* (2001) Twenty-year weathering re-measurements at St. Paul's Cathedral, London, *Earth Surface Processes and Landforms* 26, 1,129–1,142.

STEVE TRUDGILL

MICROATOLL

Microatolls are discoid colonies of massive corals that grow in the lower intertidal zone on shallow reef flats. Corals are organisms that secrete a calcareous exoskeleton and they are major contributors to a CORAL REEF that may be preserved in the geological record as limestone. Microatolls are corals that grow at the reef-atmosphere interface, adopting a predominantly lateral growth form, constrained in their upward growth by exposure at lowest tides. This upper limit to coral growth reflects physiological factors that inhibit coral polyps when exposed.

Microatolls can be formed by several species of corals, but massive corals, such as *Porites*, are particularly prominent and more likely to be preserved (Plate 80). These growth forms have received particular attention because their flat upper surface is limited in terms of upward growth by exposure and is therefore related to regional sea level: There was initially debate as to whether the distinct form of microatolls might be due to sediment accumulation on top of the coral or nutrient limitation, but it has been demonstrated that it is water level that constrains vertical coral growth (Stoddart and Scoffin 1979). It is possible to examine the pattern of growth because corals form growth bands which are generally annual and which can be detected using X-radiography. The banding within microatolls confirms that growth has been primarily lateral and can indicate periods during which the limit to coral growth has been temporarily raised or lowered, preserved as undulations on the upper surface of the colony (Plate 81).

Plate 80 Microatolls on the reef flat in the Cocos (Keeling) Islands, Indian Ocean. The two large massive corals in the foreground comprise colonies of *Porites* that are no longer living on top, but which have live polyps confined to the margin. Their upper surface is constrained by exposure during lowest tides, and the concentric rim around the margin records a time at which this water level was slightly higher. The microatolls are approximately a metre in diameter and have been growing for several decades

Microatolls have been particularly important in the reconstruction of mid and late Holocene SEA-LEVEL change. The significance of microatolls was recognized during the 1973 Royal Society expedition to the Great Barrier Reef, northeastern Australia. Microatolls were surveyed on the surface of several different reefs and their growth form was used to indicate that sea level during the mid-Holocene had been higher than present (Scoffin and Stoddart 1978). Surveying of sequences of microatolls across transects along the mainland shore of Queensland, and cross-correlation using radiocarbon dating, provided convincing evidence that relative sea level had been above present level by more than a metre around 6,000 years ago, and demonstrated that it had undergone a smooth fall since that time (Chappell 1983). Much of the Indo-west Pacific reef province has experienced relative sea levels in mid and late Holocene that have been slightly above present. Microatolls have sometimes been preserved at a height presently above that of their living counterparts on reef flats in the eastern Indian Ocean, Southeast Asia, northern Australia, and across much of the equatorial Pacific Ocean, and provide evidence of sea-level change particularly on atolls (Smithers and Woodroffe 2000).

However, it is also clear that the banding structure of microatolls can preserve details of other events in the life history of the coral. In areas where storms are experienced, overturning of the colony can occur during individual storms, or microatolls may have responded to the moating of water that can occur behind boulder ramparts formed as a result of storms. In these cases their upper surface may record elevation of water level within impounded moats above that of regional

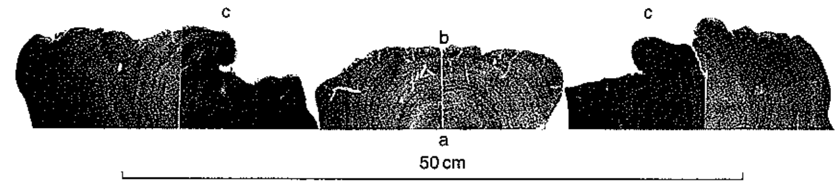

Plate 81 An X-radiograph of a vertical slice through a *Porites* microatoll. The banding indicates the growth form of the coral as the colony has aged. Initially the coral was hemispherical (a), but when it reached a level at which it was exposed too frequently during the lowest tides (b) it ceased upward growth and has extended laterally. Undulations on the surface, which occur symmetrically about the centre of the coral (c), record periods during which this upper limit to coral growth has been slightly higher

sea level. Although in open-water situations microatolls can enable centimetre-scale reconstructions of former sea level, they are subject to misinterpretation if moating has occurred, and it is therefore important to assess the geomorphological setting within which these corals grew before using fossil specimens to draw conclusions about past sea levels.

Fossil microatolls are often more accessible across the Indo-west Pacific reef province than massive hemispherical corals which are likely to have become buried by subsequent reef growth. Microatolls can be sampled along their horizontal growth axis in order to reconstruct a palaeoclimatological proxy record of the chemistry of surface waters during mid and late Holocene. Within living microatolls from the central Pacific Ocean interannual variations in water level indicate patterns of sea-level variation associated with EL NIÑO EFFECTS. Modern microatolls also enable oxygen isotope analyses of their skeleton preserving an important proxy of sea surface temperature which varies in association with the El Niño-related oscillations of sea level (Woodroffe and Gagan 2000). The application of these techniques to fossil microatolls offers the prospect of an insight into the palaeoclimatology of surface waters of more extensive areas within the tropical Indian and Pacific Oceans.

References

Chappell, J. (1983) Evidence for smoothly falling sea levels relative to north Queensland, Australia, during the past 6000 years, *Nature* 302, 406–408.
Scoffin, T.P. and Stoddart, D.R. (1978) The nature and significance of micro atolls, *Philosophical Transactions of the Royal Society of London* B284, 99–122.
Smithers, S.G. and Woodroffe, C.D. (2000) Microatolls as sea-level indicators on a mid-ocean atoll, *Marine Geology* 168, 61–78.
Stoddart, D.R. and Scoffin, T.P. (1979) Microatolls: review of form, origin and terminology, *Atoll Research Bulletin* 224, 1–17.
Woodroffe, C.D. and Gagan, M.K. (2000) Coral microatolls from the central Pacific record late Holocene El Niño, *Geophysical Research Letters* 27, 1,511–1,514.

SEE ALSO: atoll; reef

COLIN WOODROFFE

MICROMORPHOLOGY

The investigation of sediment and soil sequences has relied heavily on characterization of field morphological properties supplemented by data derived from laboratory analysis. Micromorphology, also called micropedology, extends this approach to the microscopic scale through the collection and analysis of undisturbed samples. The manufacture of soil thin sections involves the drying of samples (by air and oven drying or acetone replacement or freeze-drying), impregnation of resin under reduced pressure and curing of blocks, and sawing, lapping and mounting on glass slides. The final results are thin sections c.30 μm thick, which can be examined using a petrological microscope. It is possible to examine in thin sections the nature and spatial arrangement of such features as rock fragments, individual mineral grains, weathering products, void space, coatings of clay, organic material, precipitation of carbonate and excrement from soil animals.

W.L. Kubiena is widely regarded as the father figure to soil micromorphology and is best known for his book *The Soils of Europe* (Kubiena 1953). He very much established micromorphology's contribution to soil classification and our understanding of soil formation processes. Other benchmark contributions have been by Brewer and Sleeman (1988) in terms of the first attempt to systematize terminology, FitzPatrick (1993) for pedological interpretation and Bullock *et al.* (1985) for devising an international description system. The provision of full descriptions is extremely instructive since the investigator has to systematically examine all thin section attributes. Basic concepts from the international descriptive system are summarized in Tables 28 and 29. The application of micromorphology can be illustrated by summarizing some examples from soil science, archaeology and Quaternary science.

The application of micromorphology and its associated techniques provide distinctive insights into soil processes as induced by physical, biological and land use mechanisms (Miedema 1997). It has been used to determine the effects of different cultivation techniques on soils (Drees *et al.* 1994) and the interactions of soil and the soil biota (Kooistra 1991). As one example, Davidson *et al.* (2002) investigated the impacts of fauna on an upland grassland soil using micromorphological analysis. The incidence of eight different types of excrement in upper soil horizons was quantified using point counting. In both the organic horizon (H) and the underlying organo-mineral horizon (Ah), the bulk of the soil volume consisted of excrement derived primarily from enchytraeids and earthworms. Micromorphological analysis was thus able to demonstrate that the organic matter in these horizons was primarily derived from a limited range of soil fauna.

The early work on micromorphology as applied to archaeology very much focused on the palaeoenvironmental interpretation of buried soils as a means of providing environmental contexts for archaeological sites. This remains a key concern, but increasing attention is being given to site taphonomy and the wider impact of anthropogenic activity on soil landscapes. Courty *et al.* (1994) highlight the ways by which micromorphology can assist with understanding the relationships between environment and human behaviour. They summarize intra-site analysis (anthropogenically structural transformations, animal effects, anthropogenic deposits and spatial and temporal variability) and off-site analysis (land use practices and human-induced soil alteration and effects on landscape dynamics). As an example, they list micromorphological features and fabrics associated with such land use practices as slash and burn, up-rooting, ploughing, manuring, irrigation, horticultural practices and pasturing-herding. Thus ploughing, for example, results in fragmented slaking crusts, dusty silty clay intercalations and coatings, mixing of horizons, loss of the fine fraction, decrease in biological activity and changes in biological fabrics.

Micromorphology has contributed to Quaternary science, primarily through the investigation of buried soils. As an example, Zárate *et al.* (2000) demonstrate the particular contribution that micromorphological analysis can make to the investigation of a 5 m Holocene alluvial record in Argentina. Changes in the ratio of shells:diatoms:non-bioclastic coarse materials (e.g. quartz, feldspars) are used to propose different modes of deposition over time, for example the upper aeolian and alluvial units are dominated by non-bioclastic coarse particles. Two palaeosols are distinguished in thin sections by the presence of black isotropic, partially degraded root fragments, derived from the original vegetation cover. Microstructural changes down the section indicate the extent of pedogenic modification of the sedimentary fabric. The presence of partially welded excrements in the palaeosols indicates the former effect of soil fauna. Overall, the field micromorphological data provide the basis to a pedosedimentary reconstruction for the Holocene in this site in the Argentinian pampas. An alternative approach is to investigate soil formation on surfaces of different age. Srivastava and Parkash (2002) demonstrate the polygenetic nature of soils on the Gangetic Plains through micromorphological analysis of samples collected from surfaces ranging in age from 135,000 to < 500 BP. Of particular interest was the degradation of early clay pedofeatures by bleaching, loss of preferred orientation and the development of a coarse speckled appearance and fragmentation; in contrast, clay pedofeatures from more recent soils were thick, smooth, strongly birefringent and microlaminated.

Kemp (1998) overviews the contribution of micromorphology to palaeopedological research and he stresses the importance of relying on a combination of such features as channels, faunal excrements, calcitic root pseudomorphs, and

Table 28 Terminology in micromorphology

Soil fabric
Total organization of a soil as expressed by the spatial arrangements of the soil constituents (solids and voids)

Soil structure
Size, shape and arrangement of primary particles and voids in both aggregated and nonaggregated material and size, shape and arrangement of any aggregates present

Soil microstructures
Structures evident at magnifications $>\times 5$

Coarse and fine material
A division between coarse and fine material – division e.g. at 10 μm or 2 μm

Basic components
These are the simplest mineral and organic particles seen in thin section. They form the building blocks to the soil organization

Groundmass
General term to describe the coarse and fine material which forms the base material of soil

Micromass
General term for finer material

Pedofeature
A distinct fabric unit; stands out in contrast to adjacent soil material, e.g. clay coating in void, nodules of iron oxides

Table 29 Descriptive framework for micromorphology

1 *Structure*
2 *Groundmass*
Subdivided into coarse material and micromass
3 *Organic components*
(a) plant residues (>5 cells connected in original tissues)
(b) organic fine material (<5 cells and includes amorphous components)
(c) organic pigment staining – whole or part of micromass
Excrement pedofeatures can also be described under this heading
4 *Textural concentration features*
Features associated with an increase in concentration of material of a particular size, e.g. coatings, infillings, cappings
5 *Amorphous concentration features*
Appear amorphous in plane polarized light (PPL), isotropic in cross polarized light (XPL). Three types identified under oblique incident light are:
(a) white or dark brown colours – organic components
(b) black to yellowish brown – oxides and hydroxides of manganese
(c) yellowish brown to reddish brown – oxides and hydroxides of iron
These occur as e.g. nodules, segregations (e.g. mottles) which are impregnative
6 *Crystalline concentration features*
E.g. features consisting of Fe oxides, gypsum, gibbsite

illuvial clay coatings. He highlights fundamental challenges posed by the polygenetic nature of soils and the fact that in contemporary soils, attributes may not be in equilibrium with current environmental conditions. This leads him to discuss the equifinality problems – the same end result can come from varying combinations of processes. He illustrates this with reference to argillic horizons; in Quaternary studies the traditional view is to regard the accumulation of illuvial clay as occurring under temperate (interglacial), seasonally dry climates under stable forest cover. Kemp (1998) argues that such illuviation can occur under a range of environmental conditions and thus it would be erroneous to propose a particular palaeoenvironmental condition because of evidence of translocated clay.

In summary, micromorphological analysis can yield distinctive information on past processes of soil and sediment formation. The continuing development of image analytical techniques is encouraging more quantitative approaches. However, research using micromorphological analysis is usually combined with other approaches, for example, soil physical and chemical analysis, pollen analysis, magnetic susceptibility, organic geochemistry or microprobe analysis.

References

Brewer, R. and Sleeman, J.R. (1988) *Soil Structure and Fabric*, East Melbourne: CSIRO Publications.
Bullock, P., Fedoroff, N., Jongerius, A., Stoops, G., Tursina, T. and Babel, U. (1985) *Handbook for Soil Thin Section Description*, Wolverhampton: Waine Research Publications.
Courty, M.A., Goldberg, P. and Macphail, R.I., (1994) Ancient people – lifestyles and cultural patterns, *Proceedings 15th World Congress of Soil Science*, 250–269.
Davidson, D.A., Bruneau, P.M.C., Grieve, I.C. and Young, I.M. (2002) Impacts of fauna on an upland grassland soil as determined by micromorphological analysis, *Applied Soil Ecology* 20, 133–143.
Drees, L.R., Karathanasis, A.D., Wilding, L.P. and Blevins, R.L. (1994) Micromorphological characteristics of long-term no-till and conventionally tilled soils, *Soil Science Society of America Journal* 58, 508–517.
FitzPatrick, E.A. (1993) *Soil Microscopy and Micromorphology*, Chichester: Wiley.
Kemp, R.A. (1998) Role of micromorphology in paleopedological research, *Quaternary International* 51/52, 133–141.
Kooistra, M.J. (1991) A micromorphological approach to the interactions between soil structure and soil biota, *Agriculture, Ecosystems and Environment* 34(1–4), 315–328.
Kubiena, W.L. (1953) *The Soils of Europe*, London: Murphy.
Miedema, R. (1997) Applications of micromorphology of relevance to agronomy, *Advances in Agronomy* 59, 119–169.
Srivastava, P. and Parkash, B. (2002) Polygenetic soils of the north-central part of the Gangetic Plains: a micromorphological approach, *Catena* 46, 243–259.
Zárate, M., Kemp, R.A., Espinosa, M. and Ferrero, L. (2000) Pedosedimentary and palaeo-environmental significance of a Holocene alluvial sequence in the southern Pampas, Argentina, *Holocene* 10, 481–488.

DONALD A. DAVIDSON

MILITARY GEOMORPHOLOGY

Military geomorphology refers to the application of geomorphic concepts, principles and technologies to military operations. This subfield links geomorphology and military science. Traditionally, military geomorphology is viewed from a perspective of the powerful influences terrain morphology has on military operations (see Winters 1998). Comparatively little attention is given to the profound effects armed conflict has on the physical landscape.

Warfare causes rapid and widespread terrain alteration. Physical landscape modification by munitions, intense vehicular movement, construction of obstacles and fortifications, and deliberate destruction are common consequences of war. Some of these activities leave erosional or depositional landforms similar to natural processes (see EQUIFINALITY) (Table 30).

Munitions and vehicle manoeuvers can alter the upper soil profile, destroy vegetation and change natural drainage patterns. These effects may persist for decades. The clay-rich landscape near Verdun, France for example, remains a pockmarked anthropogenic surface resembling GILGAI, CRATERS, DRUMLINS and HUMMOCKS nearly a century after artillery pounded the terrain during the First World War. Military defensive structures including castle moats, tank ditches, trenches and bunkers remain on the physical landscape long after their military usefulness is past. CAVES built or modified for military purposes such as in the Tora Bora region of Afghanistan or Gibraltar, remain an intricate part of contemporary landscape morphology.

Deliberate destruction of terrain is not uncommon in war. Governments in conflict, including the Russians in 1812 and 1941–1942, and the United States Union Army during the American Civil War, employed 'scorched earth' tactics. Crops, vegetation and structures were purposefully destroyed to deny the enemy their use, generating changes to erosion and deposition patterns across large areas. In the Second World War, Allied forces destroyed dams on the River Ruhr in Germany, drastically changing downstream fluvial morphology. Millions of gallons of defoliants devastated tropical rainforest and cropland soils during the Vietnam conflict. In 1991, retreating Iraqi forces set fire to over 730 oil wells in Kuwait, creating oil lakes and a durable 'tarcrete' surface of petroleum sludge, an artificial DURICRUST.

Table 30 Some possible geomorphic effects of military operations

Military activity	Possible geomorphic effects	Example
Vehicular movement	Causes COMPACTION OF SOILS, decreases soil infiltration rates, destroys vegetation and changes erosion and deposition patterns. Forms RILLS, gullies, WADIS, ARROYOS, and track scars.	Tank track scars are preserved in desert pavement of the southwestern United States over 60 years.
Use of artillery, bombs, minefields	Creates blast craters, alteration of the upper soil profile, and destruction of vegetation with subsequent changes to erosion and deposition patterns. Mined areas may remain relatively unchanged by further anthropogenic alteration.	Over 20 million blast craters were produced during the Vietnam conflict. Over 100 million mines remain from conflict in more than 90 countries worldwide.
Construction of bunkers, trenches, defensive fortifications	May form mounds, gullies, wadis, arroyos, moats, CAVES and canals.	Mounds marking defensive fortifications remain along the Normandy coast of France from the Second World War.

Reference

Winters, H.A. (1998) *Battling the Elements: Weather and Terrain in the Conduct of War*, Baltimore: Johns Hopkins University Press.

SEE ALSO: anthropogeomorphology

DANIEL A. GILEWITCH

MIMA MOUND

Also called prairie mounds and pimple mounds, Mima mounds take their name from Mima Prairie, Thurston County, Washington, USA. Such mounds are characteristically up to around 2 m in height, 25–50 m in diameter, and occur at a density of 50–100 or more to the hectare. There are many hypotheses for their origin (Cox and Gakahu 1986), including that they are erosional residuals, result from depositional processes around vegetational clumps, are the product of frost sorting, have been formed by communal rodents, are degraded termitaria, or have been created by seismic activity or groundwater vortices (Reider *et al.* 1996).

They are found from the Gulf Coast to Alberta, while in southern Africa, where they are termed *Heuweltjies*, they are widely distributed in the drier, western parts (Lovegrove and Siegfried 1989). Similar forms are also known from Argentina and Kenya. These mounds probably have many different origins, but the role of such beasts as mole rats, prairie dogs and gophers should not be underestimated.

References

Cox, G.W. and Gakahu, C.G. (1986) A latitudinal test of the fossorial rodent hypothesis of Mima mound origin, *Zeitschrift für Geomorphologie NF* 30, 485–501.
Lovegrove, B.G. and Siegfried, W.R. (1989) Spacing and origin(s) of Mima-like earth mounds in the Cape Province of South Africa, *South African Journal of Science* 85(2), 108–122.

Reider, R.G., Huss, J.M. and Miller, T.W. (1996) A groundwater vortex hypothesis for Mima-like mounds, Laramie Basin, Wyoming, *Geomorphology* 16, 295–317.

A.S. GOUDIE

MINERAL MAGNETICS IN GEOMORPHOLOGY

Mineral magnetic (or environmental magnetic) analysis provides a means of characterizing soil, sediment or rock samples on the basis of their magnetic properties. As demonstrated by rock magnetic research (Dunlop and Özdemir 1997), soils, unconsolidated sediments and solid rocks all display magnetic properties. These properties can be quantified to reveal information about the types of magnetic minerals present, their concentration and, in some circumstances, their magnetic grain size. The method shares many of the underlying principles of other methods of material characterization such as size, shape, colour, mineralogical or chemical composition and, for the geomorphologist, therefore offers a similar range of potential applications.

The magnetic behaviour of materials can be broadly classified into three types: diamagnetic (i.e. quartz, feldspar, water), paramagnetic (i.e. olivine, pyroxene, biotite) and ferromagnetic (i.e. magnetite, haematite). The first two types are relatively weak phenomena. While such materials produce a measurable magnetic response in the presence of an artificial external magnetic field (due to changes induced in electron motions within the constituent atoms), they are not capable of holding a remanence of that field. Thus, when the field is removed, the electron motions return to their previous behaviour and their magnetic properties cancel, resulting in no net spontaneous magnetization in the material.

In contrast, ferromagnetic materials display much stronger magnetic response in the presence of a magnetic field and can, in some circumstances, retain a memory (remanence) of that field after it is removed (spontaneous magnetization). Ferromagnetic behaviour can, in turn, be classified into sub-types. Of most interest in a geomorphological context are ferrimagnetic (i.e. magnetite, maghemite, greigite) and imperfect (canted) antiferromagnetic (i.e. haematite, goethite) types. These two groups show contrasting magnetic behaviour when subjected to specific laboratory measurements, although their magnetic response also varies with their exact composition and grain size. In addition, many natural samples will contain assemblages of mixed mineral types, concentrations and grain sizes, resulting in potentially complex 'bulk' magnetic behaviour.

Routine measurements (Table 31) are made at room temperature and provide (a) an insight into the magnetic mineralogy, concentration and grain size within a sample and (b) a characterization (or fingerprint) of the sample. In more advanced studies, temperature dependent magnetic properties may also be measured as these may provide more conclusive evidence in terms of (a) above. The advantages of mineral magnetic techniques include the ease of measurement, the ability to process large numbers of samples relatively quickly, their non-destructive nature (for room and low temperature measurements at least) and relatively low cost. However, the most significant advantage is the sensitivity of the instrumentation. In the majority of cases, differences in iron oxide concentrations can be detected that are well below the resolution of other methods such as X-ray diffraction, differential thermal analysis or differential chemical extractions.

Mineral magnetic analysis has a range of applications in studies of the environment (Table 32). In a geomorphological context, two such applications have received particular attention: (1) studies of WEATHERING/pedogenesis and (2) sediment tracing. The alteration and redistribution of iron that takes place within the soil environment is often reflected in corresponding changes in magnetic properties and certain soil types may show diagnostic variations in their magnetic signature with depth (Maher 1986). A common phenomena is the 'enhancement' of topsoil magnetic properties (e.g. higher values than the subsoil), where fine-grained magnetite and maghemite are formed as a result of biogeochemical transformations of iron weathered from other minerals (Dearing 2000). Like other changes induced by weathering and/or pedogenesis, all other things being equal, the degree of alteration of the REGOLITH may increase with time (although not necessarily in a linear fashion). In some circumstances therefore, the level of magnetic alteration may be used as a relative dating method (e.g. White and Walden 1994; Walden and Ballantyne 2002).

Differentiation of topsoil and subsoil materials on the basis of their magnetic properties has been widely used in sediment tracing studies within lake and fluvial sediment systems (e.g. Dearing *et al.* 1985; Dearing 2000) and at a catchment scale, soils based upon different parent lithologies can also be distinguished. Considerable potential also exists for magnetic properties to be used in studies of SOIL EROSION and redistribution on hillslope systems, where they may complement other methods such as artificial radionuclides. Despite the advantages of the method, the user must also be aware of the underlying assumptions and potential problems. Two key issues are (1) the ability to identify and fully characterize the variability of the potential sediment source types and (2) the validity of assuming that the magnetic properties remain unaltered during sediment transport and deposition (Dearing 2000).

Table 31 Common room-temperature mineral magnetic parameters and their basic interpretation

Parameter	Interpretation
χ ($10^{-6}\,m^3\,kg^{-1}$)	Initial low field mass specific magnetic susceptibility. This is measured within a small magnetic field and is reversible (no remanence is induced). Its value is roughly proportional to the concentration of ferrimagnetic minerals within the sample.
χ_{fd} ($10^{-6}\,m^3\,kg^{-1}$)	Frequency dependent susceptibility. This parameter measures the variation of magnetic susceptibility with the frequency of the applied alternating magnetic field. Its value is proportional to the amount of magnetic grains whose size means they lie at the stable single domain/superparamagnetic (<0.1 μm) boundary.
χ_{ARM} ($10^{-6}\,m^3\,kg^{-1}$)	Anhysteretic Remanent Magnetization (ARM) is proportional to the concentration of ferrimagnetic grains in the 0.02 to 0.4 μm (stable single domain) size range. The final result can be expressed as mass specific ARM per unit of the steady field applied (χ_{ARM}).
SIRM ($10^{-6}\,Am^2\,kg^{-1}$)	Saturation Isothermal Remanent Magnetization (SIRM) is the highest amount of magnetic remanence that can be produced in a sample by applying a large magnetic field. The value of SIRM is related to concentrations of all remanence-carrying minerals in the sample but is dependent upon the assemblage of mineral types and their magnetic grain size.
Soft IRM ($10^{-6}\,Am^2\,kg^{-1}$)	The amount of remanence acquired by a sample after experiencing an applied field of 40 mT. At such low fields, the high coercivity, canted antiferromagnetic minerals such as haematite or goethite are unlikely to contribute to the IRM, even at fine grain sizes. The value is therefore approximately proportional to the concentration of the low coercivity, ferrimagnetic minerals (e.g. magnetite) within the sample, although also grain-size dependent.
Hard IRM ($10^{-6}\,Am^2\,kg^{-1}$)	The amount of remanence acquired in a sample beyond an applied field of 300 mT. At fields of 300 mT, the majority of ferrimagnetic minerals will already have saturated and the value is therefore approximately proportional to the concentration of canted antiferromagnetic minerals within the sample.
IRM backfield ratios	Various magnetization parameters can be obtained by applying one or more magnetic 'reverse' or 'backfields' to an already saturated sample. The magnetization at each backfield can be expressed as a ratio of IRM_{field}/SIRM and can discriminate between ferrimagnetic and canted antiferromagnetic mineral types.

Sources: After Thompson and Oldfield (1986); Maher and Thompson (1999); Walden *et al.* (1999)

References

Dearing, J.A. (2000) Natural magnetic tracers in fluvial geomorphology, in I.D.L. Foster (ed.) *Tracers in Geomorphology*, 57–82, Chichester: Wiley.

Table 32 Environmental applications of mineral magnetic analysis

Application	Sedimentary environment	Example
Sediment correlation	Lacustrine, glacial, loess, fluvial, marine, soil erosion, etc.	Correlation of Heinrich layers in N. Atlantic sediments.
Tracing sediment provenance	Lacustrine, glacial, loess, fluvial, marine, soil erosion, etc.	Source areas of glacial sediment sequences. Soil erosion into river/lake sediment systems. Catchment fire histories.
Weathering/soil forming processes	Contemporary soils, colluvial deposits, PALAEOSOL identification, ALLUVIAL FAN surfaces.	Relative dating of weathered/pedogenic surfaces. Quaternary climate change record in Chinese LOESS sediments.
Artificial tagging of sediment	Fluvial, estuarine.	Tracing movement of fluvial sediments.
Pollution monitoring	Recent organic sediments, urban drainage, atmospheric pollution.	Industrial emissions from coal-fired power stations.

Sources: After Thompson and Oldfield (1986); Maher and Thompson (1999); Walden *et al.* (1999); Dearing (2000)

Dearing, J.A., Maher, B.A. and Oldfield, F. (1985) Geomorphological linkages between soils and sediments: the role of magnetic measurements, in K. Richards, R.R. Arnett and S. Ellis (eds) *Geomorphology and Soils*, 245–266, London: George Allen and Unwin.
Dunlop, D.J. and Özdemir, Ö. (1997) *Rock Magnetism: Fundamentals and Frontiers*, Cambridge: Cambridge University Press.
Maher, B.A. (1986) Characterization of soils by mineral magnetic measurements, *Physics of Earth and Planetary Interiors* 42, 76–92.
Maher, B.A. and Thompson, R. (eds) (1999) *Quaternary Climates, Environments and Magnetism*, 81–125, Cambridge: Cambridge University Press.
Thompson, R. and Oldfield, F. (1986) *Environmental Magnetism*, London: Allen and Unwin.
Walden, J. and Ballantyne, C.K. (2002) Use of environmental magnetic measurements to validate the vertical extent of ice masses at the last glacial maximum, *Journal of Quaternary Science* 17(3), 193–200.
Walden, J., Oldfield, F. and Smith, J.P. (eds) (1999) *Environmental Magnetism: A Practical Guide*, London: Quaternary Research Association.
White, K. and Walden, J. (1994) Mineral magnetic analysis of iron oxides in arid zone soils, Tunisian Southern Atlas, in K. Pye and A.C. Millington (eds) *The Effects of Environmental Change on Geomorphic Processes and Biota in Arid and Semi-arid Regions*, 43–65, Chichester: Wiley.

SEE ALSO: soil geomorphology; tracer

JOHN WALDEN

MINING IMPACTS ON RIVERS

Mining for heavy metals such as lead, zinc, copper, gold and silver has affected river environments since the advent of metallurgy. Pollution issues associated with heavy metal extraction from alluvial sediment or directly from their original host rock are known to have had their greatest impact on river systems since the start of the industrial revolution, *c.*1800. The fate of metal pollutants is similar to the natural sediment load; it may be stored within the channel, on the floodplains or it can be transferred out of the system via the estuaries and the ocean. It is the connectivity of river, estuarine and coastal transport systems coupled to the storage capacity of fluvial environments that determines the distribution of such sediment borne metals in a catchment. A basin-wide assessment of the fate and storage of mining related metal contaminants in the north-east of England (Macklin *et al.* 2000) indicated that only a small proportion has been flushed out into the Humber estuary with the remainder being stored in and along associated river systems.

Because of the protracted residence times of heavy metals within rivers and their floodplains (from 10^1 up to 10^4 years), metal-contaminated sediment may act as major sources of future contamination. Normal channel erosion and sedimentation processes are important in redistributing such contaminants across floodplains and for moving any inchannel material downstream. Fluvial sediments are spatially and temporarily complex, reflecting changes in the frequency and magnitude of river behaviour (erosion and sedimentation) to either extrinsic (e.g. climate or land use changes) or intrinsic (e.g. threshold adjustments) impacts. The distribution of metals in a river system is similarly complex. In more homogenous fluvial environments, it is often possible to determine pre-mining, mining era and post-mining era sediments through the examination of vertical overbank sequences. Here the deepest, oldest units may contain a geochemical imprint of the catchment prior to human disturbance. Moving up through the sediment profile one might encounter changes in sediment metal values that reflect the impact of human activity in a catchment. However, several studies (e.g. Macklin *et al.* 1994; Taylor 1996) have revealed that the imprint of mining activity may not be so simply distributed in floodplain sediments. Pre- and post-industrial anthropogenic deforestation may disrupt the predicted geochemical profile of alluvium either through dilution of the contaminant signal or through the erosion and transfer of sediments from a mineralized catchment that is naturally enriched in heavy metals. Floodplain geomorphology, such as terraces, levees, back channel environments, lakes and cut-offs may also assert a major control on the distribution of metals within floodplain environments such that overbank sediment profiles may show considerable variation in heavy metal concentrations both laterally and vertically.

The transfer of metals in rivers is dominated by four major mechanisms (Lewin and Macklin 1987): (1) hydraulic sorting according to individual particle size and mineral density; (2) chemical dispersal – solution, adsorption, the formation of Fe and Mn complexes and organic uptake (bioaccumulation in plants and animals); (3) dilution with clean uncontaminated sediments; (4) loss and exchange with floodplain sediment. Chemical remobilization and dispersal remains a significant problem particularly with respect to acid mine drainage where low pH values and changes in redox (i.e. reducing) conditions can liberate co-precipitated or adsorbed metals from relatively stable mine spoil into adjacent water and sediment bodies. Changes in water and sediment chemistry can increase the solubility, mobility and bioavailability of metal-contaminated sediment and thus make them more deleterious to the surrounding environment.

The transport of mining waste in river systems may occur in one of two modes: 'passive dispersal' or 'active transformation' (Lewin and Macklin 1987). In the 'passive dispersal' of riverborne metals, the river system remains in equilibrium and the waste is transported alongside the natural sediment load such that there is no significant alteration to the channel and floodplain's morphology.

'Active transformation' is associated with a greatly increased sediment load such that it results in a major transformation of the types, rates and/or magnitudes of geomorphic processes that control the prevailing channel morphology (Miller 1997). Gilbert's (1917) seminal paper described the effects of hydraulic gold mining in the Sierra Nevada (USA) on the tributaries of the Sacramento River between 1855 and 1884. Gilbert (1917) explained how mining waste resulted in rapid rises in the elevation of tributary channel beds during and following the cessation of mining. Aggradation was subsequently replaced by incision when the supply of mining debris declined and sediment was transferred downstream in wave-like form over time. However, the rate of incision (recovery) was markedly slower than that of aggradation (James 1993).

Other impacts of mining debris include the phytotoxic damage of riparian vegetation leading to bank destabilization, changes in sediment supply to the channel zone and ultimately planform metamorphosis from a meandering to a braided channel planform. In the Tasmanian Ringarooma River basin, tin mining during the nineteenth and early twentieth centuries (Knighton 1989, 1991) caused major changes to the width and planform of the Ringarooma River. The channel bed aggraded up to 10 m and channel width increased by up to 300 per cent. The response in the Ringarooma basin was very different to that on the Sacramento River because the source of the mining debris was much more diffuse, with mining sediments being distributed along the length of the river system as opposed to having a more geographically limited and discrete point source. Although the aggradation–incision cycle was evident in the Ringarooma and progressed downstream along with the mining debris sediment wave, the spatial pattern was highly variable due to the input of debris from tributaries all along the system.

The human imprint of land use change may manifest itself in many forms within a river basin. Those related to the direct impact of mining on rivers can result in the active transformation of a channel. This may cause major and highly visible disruption to the physical structure of fluvial environments through bed level adjustments, and cross-sectional and planform changes following the release of substantial volumes of toxic materials directly into the system. Less visible impacts such as the storage of heavy metal contaminants within a riverbed and on floodplains can result in the long-term storage of contaminants long after the primary pollution activity has dissipated. These may provide a latent but potentially insidious secondary source of pollution for future adjacent agricultural and urban land uses.

References

Gilbert, G.K. (1917) Hydraulic mining debris in the Sierra Nevada, *United States Geological Survey Professional Paper* 105.

James, L.A. (1993) Sustained reworking of hydraulic mining sediment in California: GK Gilbert's sediment model reconsidered, *Zeitschrift für Geomorphologie Supplementband* 88, 49–66.

Knighton, A.D. (1989) River adjustment to changes in sediment load: the effects of tin mining on the Ringarooma River, Tasmania, 1875–1984, *Earth Surface Processes and Landforms* 14, 333–359.

——(1991) Channel bed adjustment along mine-affected rivers of northeast Tasmania, *Geomorphology* 4, 205–219.

Lewin, J. and Macklin, M.G. (1987) Metal mining and floodplain sedimentation in Britain, in V. Gardiner (ed.) *International Geomorphology 1986 Part 1*, 1009–1027, Chichester: Wiley.

Macklin, M.G., Ridgway, J., Passmore, D.G. and Rumsby, B.T. (1994) The use of overbank sediment for geochemical mapping and contamination assessment: results from selected English and Welsh floodplains, *Applied Geochemistry* 9, 689–700.

Macklin, M.G., Taylor, M.P., Hudson-Edwards, K.A. and Howard, A.J. (2000) Holocene environmental change in the Yorkshire Ouse basin and its influence on river dynamics and sediment fluxes in the coastal zone, in I. Shennan, and J. Andrews (eds) *Holocene Land–Ocean Interaction and Environmental Change around the North Sea*, 87–96, Special Publications 166, London: Geological Society.

Miller, J.R. (1997) The role of fluvial geomorphic processes in the dispersal of heavy metals from mine sites, *Journal of Geochemical Exploration* 58, 101–118.

Taylor, M.P. (1996) The variability of heavy metals in floodplain sediments: a case study from mid Wales, *Catena* 28, 71–87.

Further reading

Macklin, M.G. (1996) Fluxes and storage of sediment-associated heavy metals in floodplain systems: assessment and river basin management issues at a time of rapid environmental change, in M.G. Anderson, D.E. Walling and P.D. Bates (eds) *Floodplain Processes*, 441–460, Chichester: Wiley.

SEE ALSO: alluvium; floodplain; fluvial erosion quantification; fluvial geomorphology; sedimentation; threshold, geomorphic

MARK PATRICK TAYLOR

MIRE

Definition of the term mire is not straightforward because mires may exist along a continuum from deep water aquatic systems through to terrestrial systems, and their boundaries are not easily defined and identified. Consequently, mires reflect different origins and patterns of development; are found in widely different geographical locations and under different climatic regimes; encompass different sets of controlling forces, and reflect different stages of successional development (Hofstetter 2000).

Mires are essentially peat-accumulating landscape features which, under natural conditions, are found where the water level is continuously near the soil surface resulting in a very narrow aerobic layer. These abiotic conditions favour specific plant and animal species, mosses or micro-organisms, which are adapted to wet and often nutrient poor conditions. The distribution of, and differences in, mire type, vegetation composition and soil type are caused primarily by geology, topography and climate but the formation, persistence, size and function of mires are controlled by *hydrological processes*. The source of the water, its quantity and quality and the mechanism by which it is delivered to the mire combine to influence mire development and character, giving rise to the wide spectrum of different mire types that occur in the landscape (e.g. Gore 1983; Heathwaite and Gottlich 1993; Moore 1984). Indeed, Mitsch and Gosselink (1993) went so far as to state that: 'Hydrology is probably the single most important determinant of the establishment and maintenance of specific types of mires and mire processes.'

There is no uniquely correct way of classifying mires because: (1) they are characterized by continuous gradation of properties with variable discontinuities, (2) their formation is affected by changing climatic, geomorphological and hydrological conditions, and (3) variations in mire types occur on a variety of scales. Thus it is not surprising that terms such as *bog, fen, mire* and *moor* are used widely but often imprecisely and occasionally interchangeably! The source of this imprecision lies with the range of criteria that have been used in attempts to define and classify mires. These criteria include floristic composition, site hydrology, site topography, water chemistry and nutritional status, and peat structure. Such criteria are sometimes used separately and sometimes in combination. The International Mire Conservation Group (www.imcg.net) suggest the following priority characteristics in differentiating mire types:

1 Source of water,
2 Prevailing hydrology-geomorphology,
3 Base content (saturation) or pH,
4 Nutrient availability and C:N, and
5 Prevailing plant communities.

Some of these classification characteristics are examined in more detail below.

Mire hydrology

The one common ingredient of all natural mires is an 'excess' of water, or at least a hydrological balance adequate to create conditions in which the surface is usually waterlogged for at least part of the year. The wet conditions characteristic of mires may result from impeded drainage, high rates of water supply or both. Water supply may consist of telluric water (i.e. water that has had some contact with mineral ground such as river water, surface runoff or groundwater discharge) or meteoric water (i.e. precipitation). Hydrological relationships play a key role in mire ecosystem processes, and in determining structure and growth. Thus different mires have a characteristic hydroperiod, or seasonal pattern of water levels, that defines the rise and fall of surface and subsurface water. An important geoindicator is the water budget of a mire, which links inputs from ground water, runoff, precipitation and physical forces (wind, tides) with outputs from drainage, recharge, evaporation and transpiration. Annual or seasonal changes in the range of water levels affect visible surface biota, decay processes, accumulation rates and gas emissions. Such changes can occur in response to a range of external factors, such as fluctuations in water source (river diversions, groundwater pumping), climate or land use (forest clearing). Waters flowing out of mires are chemically distinct from inflow waters, because a range of physical and chemical reactions take place as water passes through organic materials, such as peat, causing some elements (e.g. heavy metals) to be sequestered and others (e.g. dissolved organic carbon, humic acids) to be mobilized.

Because mire vegetation is largely responsive to primary environmental factors, such as hydrological and hydrochemical factors, hydrological classifications such as the one in Table 33 for British mires (see Heathwaite 1995) often offer the most direct explanation of mire types, because form and biotic characteristics are determined by these features.

Mire morphology

It has long been recognized that differences in topographic situation and water supply

Table 33 Hydrological classification for British mires

Source of water	Extent		
	Small (<50 ha)	Medium (50–1,000 ha)	Large (>1,000 ha)
Rainfall	Parts of basin mires	Raised mires	Blanket mires
Springs	Flushes, acid valley and basin mires	Fen basins, acid valley and basin mires	Fen massif, The Fens, Somerset Levels
Floods	Narrow floodplains	Valley floodplains	Floodplain massif

mechanism profoundly influence mire type. Such considerations formed the basis of the early but long-standing systems of mire classification. Von Post and Granlund (1926) subdivided mires into three types: ombrogenous mires, developed under the exclusive influence of precipitation; topogenous mires, irrigated by telluric water which naturally collects on flat ground and topographic hollows; and soligenous mires developed on slopes and kept wet by a supply of telluric water. Moore and Bellamy (1974) used this principle to subdivide British mires into broad groups based on their physiography which they justified on the basis that mire development is mainly determined by climatic, hydrological and geochemical conditions that control vegetation communities, leading to development of a particular mire type (Table 34).

Mire development

Viewed statically, mire types reflect a water table zonation from more or less open water through to conditions where the water table is rarely if ever above the substrate surface, though rarely far below base. Viewed dynamically, mires reflect successional or hydroseral changes. In the UK, it is difficult to demonstrate different successional stages because most hydroseres began developing at about the same time in the postglacial (Flandrian) period and are often, therefore, at similar successional stages. However, palaeoreconstruction based on the stratigraphy of accumulated peat deposits can be used to demonstrate the dynamic features of mire development over time.

The natural progression of the autogenic mire succession is in the direction of increasing acidity as the growing peat surface becomes progressively isolated from the nutritional effect of ground and soil water and more dependent on rainwater nutrition. Geomorphological criteria, principally climate, topography and substrate geology, are fundamental in directing changes to this autogenic succession. Thus the zonation of topogenous mires characteristically follows the water table gradient round enclosed basins or hollows which concentrate flow and allow the accumulation of a peat where lateral water movement is impeded. Soligenous mires develop where slow lateral gravitational seepage maintains waterlogged conditions at the ground surface. Topography is still important but it is the slow percolation of water through this mire type that distinguishes their type. Water flowing through soligenous mires is typically more oxygenated relative to the stagnant conditions of topogenous mires, and consequently the rate of organic matter decomposition is higher and the depth of peat accumulation lower. Ombrogenous mires develop where precipitation is high relative to evapotranspiration. Topography, whilst important, largely acts to retard runoff from the mire, rather than to concentrate runoff to it from other areas. The dependence on atmospheric inputs alone produces mire habitats that are characteristically of low base status, and where organic matter decomposition is low and the accumulation of unhumified acid peat high relative to other mire types. In the UK, ombrogenous mires are subdivided into raised mires and blanket mires or bogs. Blanket mires develop where the ground surface is

Table 34 Physiographical classification for British mires

Soligenous mires	Moving water/flushes/springs, slow peat formation (due to O_2)
Basin mires	Found in deep hollows, e.g. kettleholes; deep peat; vegetation surface may float; limited groundwater movement
Valley mires	Characterized by water flow along valley axis; broad range in pH, nutrients and vegetation communities
Floodplain mires	Develop on flood-prone alluvium; broad range in pH, nutrients and vegetation communities
Raised mires	Characterized by a peat surface that is isolated from the regional groundwater table; ombrotrophic; domed shape
Blanket mires	Usually develop on impermeable materials in regions with high precipitation and low temperature

permanently wet, initiating peat formation on flat and gently sloping ground. Raised mires occur over a wide range of climatic conditions and may represent a late stage in the autogenic succession of topogenous mires. In lowland Britain, raised mires are recorded in basins, floodplains and at the heads of estuaries, for example Thorne Moors National Nature Reserve which forms part of the Humberhead Peatlands. They are characterized by a raised central mire area where the peat has accumulated to the extent that it becomes isolated from water feeding the mire margins to become solely dependent on rainfall inputs. Here acidification ensues, the rate of decomposition falls, peat accumulation increases, and the mire type shifts from topogenous or soligenous mires to ombrogenous types. Raised mires form characteristic shallow domes of peat where the topography is typically convex, with a gently sloping rand away from its centre towards the surrounding moat-like drainage channel or lagg surrounding the bog.

Mire hydrochemistry

In addition to hydrological controls on mire development, the base and nutrient status of the mire water supply influence the mire type. The water chemistry of mires is primarily a result of geologic setting, water balance (relative proportions of inflow, outflow and storage), quality of inflowing water, type of soils and vegetation, and human activity within or near the mire. Mires dominated by surface-water inflow and outflow reflect the chemistry of the associated rivers or lakes. Mires that receive water primarily from precipitation and lose water by way of surface-water outflows and (or) seepage to ground water tend to have lower concentrations of chemicals. Thus ombrotrophic mires are rainwater-dominated and consequently base-deficient whereas minerotrophic mires are supplied with minerals and nutrients via the mire substrate which is in turn dependent on catchment geology and drainage water quality. Minerotrophic mires may range from oligotrophic through to mesotrophic or eutrophic types depending on the quality of their source water. Thus hydrochemical mire classifications focus largely on the source and quality of water to a mire giving a range of mire types from ombrotrophic raised mires, through transitional mires to minerotrophic mires or fens (Table 35).

On the basis of floristic variation in Swedish mires, Du Rietz (1949, 1954) suggested that mires could be divided into areas fed almost exclusively by precipitation and those in which water supply was supplemented by telluric water. These early concepts are important because they broadly correspond with major habitat differences still recognized today. The term fen is largely used as a synonym for minerotrophic mires and bog to refer to ombrotrophic examples (see Wheeler and Proctor 2000). These colloquial terms are still confused however, particularly as the vegetation of bogs and fens can be very similar. The distinction between these two habitats is based on their respective water sources. Bogs obtain their water from rainfall alone and this water is essentially stagnant, at least in the lower bog layers or catotelm. The water in a fen flows, although this may happen very slowly. Joosten (1998) used base conditions and trophic status to differentiate mire types, see Table 36.

The nutrient supply to bogs is characteristically low although nitrogen may be supplemented by

Table 35 Mean values of the concentration of major ions in waters from European mires

Mire hydrochemistry		Major ions									
		pH	HCO_3	Cl	SO_4	Ca	Mg	Na	K	H	Total
Eutrophic	1	7.5	3.9	0.4	0.8	4.0	0.6	0.5	0.05	0	10.25
↓	2	6.9	2.7	0.5	1.0	3.2	0.4	0.4	0.08	0	8.28
	3	6.2	1.0	0.5	0.7	1.2	0.4	0.5	0.02	0	4.32
	4	5.6	0.4	0.5	0.5	0.7	0.2	0.5	0.04	0.01	2.85
	5	4.8	0.1	0.3	0.5	0.3	0.1	0.3	0.07	0.03	1.70
	6	4.1	0	0.4	0.4	0.2	0.1	0.3	0.04	0.14	1.58
Oligotrophic	7	3.8	0	0.3	0.3	0.1	0.1	0.2	0.04	0.16	1.20

Source: After Moore and Bellamy (1974)

Table 36 Base conditions, trophic conditions and mire types

C/N ratio	>33				<20
pH		<4.8		>6.4	
	Oligotrophic acid	Mesotrophic acid	Mesotrophic subneutral	Mesotrophic calcareous	Eutrophic
Lowland bog					
Mountain bog					
Kettlehole mire					
Percolation mire					
Surface flow mire					
Terrestrialization mire					
Spring mire					
Coastal floodplain mire					
Fluvial floodplain mire					

Source: After Joosten (1998)

atmospheric enrichment from industrial, urban and agricultural sources. Bog biodiversity is low. Typically the pH is <4.5 compared to fens where the pH range is 4.5–7.5.

Significance of mires in the landscape

In western Europe the majority of natural mires have been degraded through anthropogenic changes in hydrology, both at the regional and local scale, primarily for agricultural purposes. These measures have affected the biotic composition, the soil physical and chemical properties, the carbon and nutrient dynamic, as well as the landscape ecological functions of mires.

The regulatory function of hydrologically undisturbed mires compared to degraded mires has been neglected until recently. Natural mires act as ecotones between terrestrial and aquatic environments and are important owing to their transformation, buffer and sink qualities. For example, minerotrophic mires or fens are connected with their surrounding terrestrial areas via several hydrological pathways such as groundwater inflow, surface runoff, interflow or river water surplus. Nutrients transported with the inflowing water into such mires are transformed or accumulated by several biogeochemical processes. As a result, the nutrient concentration in the outflow can be reduced and water quality improved. Thus lowland mires are often areas of high biological productivity and diversity and mediate large and small-scale environmental processes by altering downstream catchments. For example, lowland mires can affect local hydrology by acting as a filter, sequestering and storing heavy metals and other pollutants, and serving as flood buffers and, in coastal zones, as storm defences and erosion controls. Upland mires can act as a carbon sink, storing organic carbon in waterlogged sediments. Even slowly growing peat may sequester carbon at between 0.5 and 0.7 tonnes ha^{-1} a^{-1}. Mires can also be a carbon source, when it is released via degassing during decay processes, or after drainage and cutting, as a result of oxidation or burning. Globally, upland mires have shifted over the past two centuries from sinks to sources of carbon, largely because of human exploitation.

References

Du Rietz, G.E. (1949) Huvudenheter och huvudgränser i svensk myrvegetation, *Srensk Botanisk Tidskrift* 43, 274–309.

——(1954) Die mineralbodenwasser-zeigegrenze als Grunlage einer natürlichen Zweigliederung der nord- und Mitteleuropäischen Moore, *Vegetatio* 5/6, 571–585.

Gore, A.J.P. (1983) *Ecosystems of the World, 4B: Mires, Swamp, Bog, Fen and Moor, Regional Studies*, Amsterdam: Elsevier.

Heathwaite, A.L. (1995) The hydrology of British mires, in J. Hughes and A.L. Heathwaite (eds) *Hydrology and Hydrochemistry of British Mires*, 11–20, Chichester: Wiley.

Heathwaite, A.L. and Gottlich, Kh. (1993) *Mires – Process, Exploitation and Conservation*, Chichester: Wiley.

Hofstetter, R.F. (2000) *Universal Mire Lexicon*, Greifswald: International Mire Conservation Group.

Joosten, J. (1998) *Mire Classification for Nature Conservation*, IMCG Working paper, Greifswald: International Mire Conservation Group.

Mitsch, W.J. and Gosselink, J.G. (1993) *Mires*, New York: Van Nostrand Reinhold.

Moore, P.D. (ed.) (1984) *European Mires*, London: Academic Press.

Moore, P.D. and Bellamy, D.J. (1974) *Peatlands*, London: Elek. Science.

Von Post, L. and Granlund, E. (1926) Sodra Sveriges tortillangar I, *Sver. Geol. Unders.* 19 C 335, Stockholm.

Wheeler, B.D. and Proctor, M.C.F. (2000) Ecological gradients, subdivisions and terminology of north-west European mires, *Journal of Ecology* 88, 187–203.

Key websites

RAMSAR: http://www.ramsar.org
Irish Peatland Conservation Council: http://www.ipcc.ie
Society of Mire Scientists: http://www.sws.org
International Peat Society: http://www.peatsociety.fi
British Ecological Society Mires Research Group: http://www.britishecologicalsociety.org/groups/mires/index.php

LOUISE HEATHWAITE

MOBILE BED

A fluid, such as air or water, flowing over cohesionless sediment has the ability to entrain solid particles. The bed surface becomes mobile when the shear stress applied on the particles by the flow exceeds the critical shear stress of the sediment mixture. The initiation of particle motion is a stochastic phenomenon that depends on the average fluid motions and, because the dimensions of sediment particles usually are relatively small compared to the dimensions of the flow, on the magnitude of turbulent deviations from the average (Nelson *et al.* 2001). It also depends on the position of a particle on the bed, which determines its exposure to the flow (Kirchner *et al.* 1990; Buffington *et al.* 1992), and the relative proportion of each size fraction in a mixture (Wilcock 1993). The shear stress at which particle motion is initiated in heterogeneous sediment may be approximated by Shields's relation, if the median grain size is used to characterize the entire sediment mixture (Kuhnle 1993; Buffington and Montgomery 1997).

As the threshold condition for the initiation of particle motion is approached there is an abrupt increase in the rate of sediment movement. Particle movement is neither uniform nor continuous over the bed, because turbulent sweeps, the structures responsible for particle motion, move groups of particles intermittently at random locations on the bed (Drake *et al.* 1988; Williams *et al.* 1990). As the shear stress (and rate of sediment transport) increases the sweeps become more laterally stable and longitudinal streaks form on the bed, and in heterogeneous sediment a pattern of alternating coarse- and fine-grained stripes emerges (McLelland *et al.* 1999). At higher shear stresses the coarser sediment becomes more mobile and the stripes are replaced by flow-transverse BEDFORMS (Gyr and Müller 1996).

Sediment initially was thought to move in sliding layers, with the most rapidly moving layer positioned adjacent to the flow, but it was soon recognized that only the surficial grains move. In the absence of significant scour or bedform development, the depth of the active layer is of the order of 0.4 to $2D_{90}$ (where D_{90} is the size for which 90 per cent of the surficial bed material size distribution is finer). The sediment in transport is termed the bed material load. Bed material either can be swept up into the main part of the flow by turbulence, and transported in suspension, or it may move, by rolling/sliding or SALTATION, as BEDLOAD in a layer immediately above the bed. This layer is of the order of two to four grain diameters thick in water, and a few tens of centimetres thick in air. As the flow intensity increases above the critical value, particles first move by rolling. Saltation rapidly becomes the dominant type of motion as the flow intensity increases further, and at still higher flow intensities suspension begins to dominate. There is a clear physical difference between the two basic modes of transport (Abbott and Francis 1977). The weight of a saltating particle is supported by the bed, whereas the flow supports the weight of a suspended particle. However, the two modes of transport cannot easily be differentiated on the basis of particle size, and there is a continual exchange of particles between the bedload and SUSPENDED LOAD. There is also an important difference between the movement of particles by saltation, in air and in water. In air, once saltation commences subsequent movement is induced by the impact of particles hitting the bed, rather than by the hydrodynamic forces that act on static particles, as is the case in water. The difference arises because the submerged density of sediment particles is substantially greater than the density of air at atmospheric pressure, whereas it is less than twice the density of water.

Particles transported in suspension move at the velocity of the flow. Particles comprising the bedload continually move in and out of storage on the bed, and their pattern of motion can be characterized as a series of relatively short steps of random length, each of which is followed by a rest period of random duration (Habersack 2001). The sensitivity of travel distance to particle size decreases as size decreases below the median diameter of the substrate (Church and Hassan 1992), but the virtual velocity of particles in water is only of the order of metres per hour, compared to the flow velocity which may be of the order of metres per second (Haschenburger and Church 1998). This is because each particle spends a negligible time in motion compared to the time spent at rest. In the case of sediment that is deposited on the lee side of a bedform, the velocity at which the particles move is much slower and is determined by the rate of movement of the bedform (Grigg 1970; Tsoar 1974).

References

Abbott, J.E. and Francis, J.R.D. (1977) Saltation and suspension trajectories of solid grains in a water stream, *Philosophical Transactions of the Royal Society of London* A284, 225–254.

Buffington, J.M. and Montgomery, D.R. (1997) A systematic analysis of eight decades of incipient motion studies, with special reference to gravel-bedded rivers, *Water Resources Research* 33, 1,993–2,029.

Buffington, J.M., Dietrich, W.E. and Kirchner, J.W. (1992) Friction angle measurements on a naturally formed gravel streambed: implications for critical boundary shear stress, *Water Resources Research* 28, 411–425.

Church, M.A. and Hassan, M.A. (1992) Size and distance of travel of unconstrained clasts on a streambed, *Water Resources Research* 28, 299–303.

Drake, T.S., Shreve, R.L., Dietrich, R.L., Whiting, P.J. and Leopold, L.B. (1988) Bedload transport of fine gravel observed by motion-picture photography, *Journal of Fluid Mechanics* 192, 193–217.

Grigg, N.S. (1970) Motion of single particles in alluvial channels, *Journal of the Hydraulics Division*, American Society of Civil Engineers 96, 2,501–2,518.

Gyr, A. and Müller, A. (1996) The role of coherent structures in developing bedforms during sediment transport, in P.J. Ashworth, S.J. Bennett, J.L. Best and S.J. McLelland (eds) *Coherent Flow Structures in Open Channels*, 227–235, Chichester: Wiley.

Habersack, H.M. (2001) Radio-tracking gravel particles in a large braided river in New Zealand: a field test of the stochastic theory of bed load transport proposed by Einstein, *Hydrological Processes* 15, 377–391.

Haschenburger, J.K. and Church, M.A. (1998) Bed material transport estimated from the virtual velocity of sediment, *Earth Surface Processes and Landforms* 23, 791–808.

Kirchner, J.W., Dietrich, W.E., Iseya, F. and Ikeda, H. (1990) The variability of critical shear stress, friction angle, and grain protrusion in water-worked sediments, *Sedimentology* 37, 647–672.

Kuhnle, R.A. (1993) Fluvial transport of sand and gravel mixtures with bimodal size distributions, *Sedimentary Geology* 85, 17–24.

McLelland, S.J., Ashworth, P.J., Best, J.L. and Livesey, J.R. (1999) Turbulence and secondary flow over sediment stripes in weakly bimodal bed material, *Journal of Hydraulic Engineering* 125, 463–473.

Nelson, J.M., Schmeeckle, M.W. and Shreve, R.L. (2001) Turbulence and particle entrainment, in M.P. Mosley (ed.) *Gravel-Bed Rivers* V, 221–248, New Zealand Hydrological Society, Wellington.

Tsoar, H. (1974) Desert dunes, morphology and dynamics, El Arish (northern Sinai), *Zeitscrift für Geomorphology Supplementband* 20, 41–61.

Wilcock, P.R. (1993) Critical shear stress of natural sediments, *Journal of Hydraulic Engineering* 119, 491–505.

Williams, J.J., Butterfield, G.R. and Clark, D.G. (1990) Rates of aerodynamic entrainment in a developing boundary layer, *Sedimentology* 37, 1,039–1,048.

BASIL GOMEZ

MODELS

'The sciences', wrote mathematician John von Neumann (1963), 'mainly make models.' Model building is as much a part of twenty-first-century geomorphology as it is in any science. A model, in its most general sense, is a simplified or idealized representation of an existing or potential reality. Examples of models range from architects' miniatures to quantum theory. In geomorphology, models serve as representations of Earth surface processes and landforms, and as such they embody the theory that underpins the science. All geomorphologists rely on models of one sort or another. However, the word 'model' is used in a variety of contexts in geomorphology, and it is useful to distinguish between three general forms: conceptual models, hardware (or experimental) models and mathematical models.

Conceptual models of landform origins must surely predate the word 'geomorphology'. As the formal scientific field of geomorphology began to take shape in the late nineteenth century, influential figures such as William Morris Davis, Walther Penck and G.K. Gilbert developed conceptual models of landscape systems that provided a guiding impetus for research. Davis's 'geographical cycle' is a classic example of a conceptual model in geomorphology. It provided an explanation for many observed landforms, made predictions about their course of evolution, provided guiding assumptions for the interpretation of particular landform elements (such as low-relief surfaces) and influenced the type of questions posed by researchers. Although many of Davis's ideas have not stood the test of time, the creation and progressive refinement of conceptual models are still fundamental parts of geomorphology. Unavoidably, our ideas about how geomorphic systems operate will always guide the type of questions we choose to ask and the kind of interpretations we make (Brown 1996).

Hardware models represent geomorphology's experimental side. A hardware model is a physical representation, often (but not always) scaled down, of a particular geomorphic system. G.K. Gilbert's (1914) flume experiments at Berkeley, which led to his classic paper 'The transportation of debris by running water', represent one of the first experimental studies in geomorphology. Gilbert's data are, in fact, still used today, and have been complemented by many other flume studies of sediment transport. The literature is replete with examples of experimental models of geomorphic systems. Phenomena that have been studied experimentally include drainage basin evolution, bedrock landsliding, soil creep, rock weathering, alluvial fans, and subaerial and subaqueous debris flows, to name a few. In some cases, laboratory experiments have coupled geomorphic processes with tectonic, eustatic, and/or depositional processes. In some instances, hardware models operate on the same spatial scale as the geomorphic system in question; the US Geological Survey experimental debris flow flume near Bellingham, Washington, USA is one such example (Major and Iverson 1999). More commonly, the physical system is scaled down, which can introduce problems in preserving basic scaling relationships between physical properties such as fluid viscosity and gravity. Nonetheless, hardware modelling continues to be an important source of information and insight into a wide range of geomorphic systems.

A mathematical model, like a conceptual model, acts as a simplified or idealized representation of reality that provides a framework for guiding and interpreting observations. Seen in this light, a mathematical model can be understood as a quantitative hypothesis or set of linked hypotheses. A mathematical model has an obvious advantage over purely conceptual models in its precision, lack of ambiguity and ability to satisfy basic constraints such as continuity of mass, momentum and energy. At the same time, mathematical models, like hardware models, allow for a degree of experimentation – in the sense of testing the behaviour of a system (one comprising a set of mathematical-logical postulates and assumptions, rather than a physical construct) that has been built as an analogy for a natural system. The use of ocean and atmospheric general circulation models to test palaeoclimate hypotheses (e.g. Cane and Molnar 2001) is a good example of this type of experimentation.

Although there is no generally agreed classification of types mathematical model, the categories suggested by Kirkby *et al.* (1992) provide a useful framework. They distinguish between black-box (statistical or empirical) models, process models, mass-balance models and stochastic models. There is significant overlap among these categories, and indeed many mathematical models combine elements of several of these.

Mathematical models in geomorphology began to emerge in the post-Second World War era. Many of these early models were descriptive or empirical (i.e. 'black box') in nature. R.E. Horton's drainage network laws (now known as HORTON'S LAWS), for example, provided a quantitative description of river network topology, while the hydraulic geometry equations of Leopold and Maddock (1953) provided a similar description of river channel changes through space and time. These and many other morphometric models are essentially statistical in nature.

Beginning in the 1960s, such statistical models were complemented by process models. Where a black-box model represents relationships in a purely empirical form, a process model attempts to describe the mechanisms involved in a system. For example, a black-box model of soil erosion would be based on regression equations obtained directly from data, whereas a process model would attempt to represent the mechanics of overland flow and particle detachment. Process models often overlap with mass-balance (or energy-balance) models, in the sense that equations for processes are used to model the transfer of mass or energy among different *stores*, where a store could represent anything from the water in a lake, the population of a species in an ecosystem, the energy stored as latent heat in an atmospheric column, the carbon mass in a tree, or the depth of soil at a point on a hillslope.

Process models have been widely used to study landform evolution. Usually phrased in terms of continuum mechanics, these landform evolution models provide a link between the physics and chemistry of geomorphic processes, and the shape of the resulting topography. Among the pioneers in landform 'process-response' modelling in the late 1960s and early 1970s were F. Ahnert and M.J. Kirkby. The latter showed, for example, that the convexo-concave form of hillslopes can be predicted from simple laws for sediment transport (Kirkby 1971).

Many process models are deterministic, meaning that for a given set of inputs they will predict a unique set of outputs. Often, however, the inputs to a particular geomorphic system are highly variable in time or space, and essentially unpredictable or unmeasurable. For example, we may know something about the frequency and magnitude characteristics of rainfall but cannot predict the sequence of rainfall events over time spans of more than a few days. Similarly, we may have a good estimate of the average hydraulic conductivity of an aquifer but little or no information about its heterogeneity. Likewise, a hallmark of many nonlinear systems (including some geomorphologic systems) is sensitivity to initial conditions: a small difference in the initial state of a system can lead to markedly different outcomes (see Gleick 1988). *Stochastic* models are designed to address such uncertainties by including an element of random variability. Such models typically use a random number generator to create a series of alternative inputs (e.g. rainstorms) or to trigger discrete events (e.g. landslides). Discussion and examples of stochastic models are given by Kirkby *et al.* (1992, Chapter 5).

At the heart of most geomorphic process models, both deterministic and stochastic, lies the continuity of mass equation

$$\frac{\partial \eta}{\partial t} = -\nabla q_s$$

where η is the height of the surface, t is time, q_s is the bulk volume rate of mass transport (rock, sediment or solute) per unit width and ∇ denotes a gradient in two dimensions. The continuity equation simply states mathematically that matter can neither be created nor destroyed (short of nuclear reactions). This particular form of the continuity equation is not universally applicable to all geomorphic problems; a slightly different form would be needed to describe horizontal (as opposed to vertical) retreat of a cliff face, the evolution of a fault block undergoing horizontal motion, or surface change due to changes in density rather than in mass, for example. Nonetheless, the continuity law in its various guises is arguably one of a handful of fundamental principles in geomorphological modelling. When combined with a suitable expression for q_s that represents a particular process or processes, the continuity equation can, subject to certain simplifying assumptions, be solved in order to predict landform features such as the shape of hillslope profiles, river profile geometry and soil-depth profiles.

An obvious advantage of mathematical process models over conceptual models (such as the influential concepts of G.K. Gilbert) is that they allow one to make statements not only about 'how' or 'why' but also 'how much' – for example, a mathematical slope model allows one to predict, based on processes, the degree to which the relief in a mountain drainage basin would ultimately change if the rate of uplift were to double (e.g. Snyder *et al.* 2000).

Central to geomorphic process models is the concept of 'geomorphic transport laws' (Dietrich *et al.* 2003). A geomorphic transport law is a mathematical statement about rates of mass transport averaged over a suitably long period of time. The definition of 'suitably long' depends on the process in question, but in general is much longer than the recurrence interval of discrete transport events such as floods, raindrop impacts, landslides, and so on. One of the current research frontiers in geomorphic process research lies in understanding the relationship between short-term transport events and long-term average transport rates.

Solving the continuity equation to predict landform shape requires assuming idealized conditions – for example, in the case of landforms with uniform soil or sediment properties, uniform climate, and height variations in one direction only. Modelling three-dimensional landforms generally requires approximating the solution to an appropriate form of the continuity equation, usually through the use of numerical techniques such as finite differencing, finite volume or finite element methods, cellular automata, or (in some cases) a combination of methods (e.g. Press 2002; Slingerland *et al.* 1994). Typically, these methods produce an approximate solution by dividing up space into discrete elements. Fluxes of mass are then computed within or between these elements. Beginning with a specified initial landform configuration, the evolution of landforms over time is computed by iteratively calculating the transport rates at each point, extrapolating these rates forward in time over a discrete time increment, and then adjusting the topography accordingly. This in turn affects the transport rates at the next time increment, so that the landform emerges as the result of an interaction between its shape and the processes acting upon it.

The development and use of numerical models of landform evolution has grown considerably since the 1980s. Examples include models of rill erosion, river basin evolution, glacial valley formation, and many other coupled tectonic–geomorphic–sedimentary systems. Numerical models of landform and landscape evolution generally operate on what Schumm and Lichty (1965) termed 'cyclic time' – that is, time spans on which landforms can change significantly and which, apart from rapid processes like rill erosion, are generally much longer than a human lifetime. Alongside these 'cyclic time' models are 'event time' numerical models aimed at understanding process dynamics. Here, event time refers to the timescales of individual process events such as floods. This is the timescale on which direct experimentation, observation and application of Newtonian mechanics are most feasible. For example, computational fluid dynamics models have been used to great effect to examine phenomena such as river flooding, coastal sediment transport, soil erosion and debris flows. Often, such models are founded on basic theory in fluid dynamics or material rheology, and combine well-established physical principles (e.g. the Navier–Stokes equations for fluid flow) with empirical laws obtained from laboratory experiments (for background and examples, see Middleton and Wilcock, 1994).

Both event-time and cyclic-time models have had a tremendous impact on geomorphologists' ability to understand the dynamics of processes, and to link these with the landforms that they shape. The applications of models range quite widely, and include both pragmatic forecasting and investigative analysis. Event-time models are

often used in an applied context, to make predictions for purposes of planning, land management and insurance assessment. Models of soil erosion, for example, are typically used in this way. In an applied, predictive mode of application, a given model is generally taken 'as read', usually calibrated with existing data, and used to forecast the outcomes of different scenarios.

Numerical models in geomorphology serve other important roles as well. Both event-time and cyclic-time models have been, and continue to be, used in a heuristic mode; that is, they are used as theoretical tools for developing general insight and understanding, rather than for making precise predictions in a particular case study. One of the most valuable roles of mathematical models in geomorphology, in fact, is to make testable predictions about process and form connections. For example, numerous river basin evolution models have been used in 'what if' mode to predict the morphological consequences of statements such as 'the long-term average incision rate of a stream channel is proportional to the rate of energy dissipation per unit bed area' (e.g. Whipple and Tucker 1999). This exploratory process of *forward modelling* makes it possible to reject some models in favour of others, based on their ability to reproduce observed landform characteristics given a plausible set of initial and boundary conditions.

As in other sciences, models in geomorphology both drive and are driven by the results of observational and experimental work. In some cases, a model is developed for the express purpose of explaining a set of data. In others, one or more models are proposed before any relevant data exist, and they stimulate the search for new observations. One example of the latter concerns the relationship between the thickness of soil and the rate of lowering of the soil–bedrock contact. Several models were proposed in the 1960s and 1970s (see Cox 1980). Of these, some predicted an exponential decline in regolith production rate with increasing soil depth, with the maximum production rate occurring at or near the surface. Others predicted a 'humped curve' with a maximum production rate at some optimal soil thickness, due to the added efficiency of water retention. These models remained essentially untested for many years, until cosmogenic nuclide analysis made it possible to infer rates of regolith production. Research beginning in the 1990s has provided evidence for an inverse dependence of regolith production rate on regolith thickness, in some cases with a near-surface maximum (e.g. Heimsath *et al.* 1997), in others with a maximum at depth (e.g. Small *et al.* 1999), depending on process and environment.

The example of regolith production models serves as a caution against the common myth that a model is of no value until and unless it has been validated. In fact, untested mathematical models in geomorphology – like the regolith production models when they were first proposed – have served the field well in two ways: first, by forcing rigour into our hypotheses, and second, by spurring the development of new efforts, ideas and technologies to test the models (for discussion see Bras *et al.* 2003).

A common limitation of models in geomorphology is that different models predict similar outcomes. For example, a range of different river process models predict that graded river profiles should be concave-upward in form – thereby providing multiple, competing explanations for the same observation. This classic problem of EQUIFINALITY, which is common across the Earth sciences, reflects a paucity of data about geomorphic systems. This limitation is part and parcel of the deep-time problem in the Earth sciences (and in other fields such as astronomy and astrophysics). The systems that geomorphologists study are often too big or too slow to allow for direct experiments. Furthermore, most geomorphic systems are dissipative in nature (Huggett 1988). Dissipative systems, by the 2nd law of thermodynamics, lose information as they evolve (consider, for example, trying to reconstruct a snow crystal from a drop of water). Geomorphologists are therefore forced to rely on inference, analogy and indirect evidence. It is no surprise that the problems of equifinality and deep time limit mathematical modelling in the same way that they limit geomorphic knowledge more generally. In principle, the solution to both problems is to obtain as much information as possible about denudation rates, boundary conditions (such as tectonic, climatic or sea-level variations), and the nature of changes in topography over the geologic past. Developing techniques to obtain such data arguably constitutes one of the foremost challenges in geomorphology.

The deep-time problem highlights the fact that, in building models of landform genesis, geomorphologists are forced to 'scale up' contemporary processes over geologic time, and over spatial scales relevant to the landforms in question. This approach is of course a natural outgrowth of Hutton's (1795) ideas. The impossibility of direct experiments makes it especially important to develop accurate constitutive process laws, and to pay careful attention to the role of natural variability in driving forces (such as weather and climate) and in materials (such as soil properties). This represents a considerable scaling challenge, because the formative processes often occur on timescales that are vastly smaller than the timescale required for significant landform change. For example, floods may last for minutes to days while the river basins they sculpt may take shape over hundreds of thousands of years.

Despite their limitations, the future of mathematical models in geomorphology looks bright. Continuing advances in computing power will make solving the scaling problem easier by allowing modellers to link together a wider range of time and space scales. While the deep-time problem will never go away, geomorphologists' ability to explore and test multiple working hypotheses will continue to grow. Likewise, continuing improvements in data describing the Earth's surface topography and in technologies for dating and estimating rates of change will make it possible to test models with increasing degrees of precision.

References

Bras, R.L., Tucker, G.E. and Teles, V. (2003) Six myths about mathematical modeling in geomorphology, in P.R. Wilcock and R. Iverson (eds) *Prediction in Geomorphology*, Geophysical Monograph, Washington, DC: American Geophysical Union.

Brown, H.I. (1996) The methodological roles of theory in science, in B.L. Rhodes and C. Thorn (eds) *The Scientific Nature of Geomorphology*, 3–20, Binghamton, NY: State University of New York.

Cane, M.A. and Molnar, P. (2001) Closing of the Indonesian seaway as a precursor to East African aridification around 3–4 million years ago, *Nature* 411(6,834), 157–162.

Cox, N.J. (1980) On the relationship between bedrock lowering and regolith thickness, *Earth Surface Processes and Landforms* 5(3), 271–274.

Dietrich, W.E., Bellugi, D., Sklar, L., Stock, J.D., Heimsath, A.M. and Roering, J.J. (2003) Geomorphic transport laws for predicting landscape form and dynamics, in P.R. Wilcock and R. Iverson (eds) *Prediction in Geomorphology*, Washington, DC: American Geophysical Union, 103–132.

Gilbert, G.K. (1914) The transportation of debris by running water, *US Geological Survey Professional Paper* 86.

Gleick, J. (1988) *Chaos: Making a New Science*, London: Heinemann.

Heimsath, A.M., Dietrich, W.E., Nishiizumi, K. and Finkel, R.C. (1997) The soil production function and landscape equilibrium, *Nature* 388(6,640), 358–361.

Huggett, R.J. (1988) Dissipative systems; implications for geomorphology, *Earth Surface Processes and Landforms* 13(1), 45–49.

Hutton, J. (1795) *Theory of the Earth*, Edinburgh.

Kirkby, M.J. (1971) Hillslope process-response models based on the continuity equation, slopes, form and process, *Transactions of the Institute of British Geographers, Special Publication* 3, 15–30.

Kirkby, M.J., Naden, P.S., Burt, T.P. and Butcher, D.P. (1992) *Computer Simulation in Physical Geography*, Chichester: Wiley.

Leopold, L.B. and Maddock, T., Jr (1953) The hydraulic geometry of stream channels and some physiographic implications, river morphology, *US Geological Survey Professional Paper* 252.

Major, J.J. and Iverson, R.M. (1999) Debris-flow deposition; effects of pore-fluid pressure and friction concentrated at flow margins, *Geological Society of America Bulletin* 111(10), 1,424–1,434.

Middleton, G.V. and Wilcock, P.R. (1994) *Mechanics in the Earth and Environmental Sciences*, Cambridge: Cambridge University Press.

Press, W.H. (2002) *Numerical Recipes in C++ : The Art of Scientific Computing*, Cambridge: Cambridge University Press.

Schumm, S.A. and Lichty, R.W. (1965) Time, space, and causality in geomorphology, *American Journal of Science* 263(2), 110–119.

Slingerland, R., Furlong, K. and Harbaugh, J.W. (1994) *Simulating Clastic Sedimentary Basins*, Englewood Cliffs, NJ: PTR Prentice Hall; Prentice-Hall International.

Small, E.E., Anderson, R.S. and Hancock, G.S. (1999) Estimates of the rate of regolith production using ^{10}Be and ^{26}Al from an alpine hillslope, *Geomorphology* 27(1–2), 131–150.

Snyder, N.P., Whipple, K.X., Tucker, G.E. and Merritts, D.J. (2000) Landscape response to tectonic forcing; digital elevation model analysis of stream profiles in the Mendocino triple junction region, Northern California, *Geological Society of America Bulletin* 112(8), 1,250–1,263.

von Neumann, J. (1963) The role of mathematics in the sciences and in society, and method in physical sciences, in *J. von Neumann – Collected Works Vol. VI*, ed. A.H. Taub, 477–498, New York: Macmillan.

Whipple, K.X. and Tucker, G.E. (1999) Dynamics of the stream-power river incision model; implications for height limits of mountain ranges, landscape response timescales, and research needs, *Journal of Geophysical Research*, B, Solid Earth and Planets 104(8), 17,661–17,674.

Further reading

Harmon, R.S. and Doe, W.W. III (eds) (2001) *Landscape Erosion and Evolution Modelling*, New York: Kluwer Academic/Plenum Publishers.

Rhodes, B.L. and Thorn, C.E. (ed.) (1996) *The Scientific Nature of Geomorphology*, Chichester: Wiley.

Wilcock, P.R. and Iverson, R. (ed.) (2003) *Prediction in Geomorphology, Geophysical Monograph*, Washington, DC: American Geophysical Union.

SEE ALSO: complexity in geomorphology; computational fluid dynamics; equifinality; laws, geomorphological; mathematics; mechanics of geological materials; non-linear dynamics

GREG TUCKER

MORAINE

A moraine is a glacial landform created by the deposition or deformation of sediment by glacier ice. Many different types of moraine exist, reflecting the many different processes by which glaciers deposit and deform sediment and the many locations and environments within the glacier system where deposition can occur. The material of which moraines are composed, which is generally referred to as till, is also highly variable, as its characteristics depend on the characteristics of the debris supplied by the glacier as well as on the processes and environment of GLACIAL DEPOSITION.

The term moraine has been used in a variety of different ways since it was originally introduced, and its definition remains controversial. Swiss naturalist Horace-Bénédict de Saussure originally introduced the term in 1779, and recognized that ancient moraines represented former extensions of existing glaciers. For the next two centuries the term was widely used to describe landforms created by glacial deposition, the sedimentary material of which the landforms were composed, and the debris in transport within, beneath or on the surface of glaciers. Although modern geomorphological definitions limit the term specifically to landforms, it is still sometimes applied more widely to glacial debris and glacially derived sediment. Many of the compound expressions that feature the term moraine, such as ground moraine and medial moraine, conflate elements of these different definitions, and so moraine continues to be used ambiguously in some geomorphological, glaciological and sedimentological literature. Dreimanis (1989) provides a useful review of the history of the term.

Moraines are classified both genetically according to the process by which they are created and geographically according to their position within the glacier system. There is a fundamental distinction between moraines that occur on the surface of the ice and moraines that occur on the ground surface beneath or at the margin of a glacier. Moraines on the ice surface, known as supraglacial moraines, are ephemeral features that move with the ice and are likely to be destroyed or redeposited on the ground surface when the ice beneath them ablates. They are not true landforms, and inherit the term moraine from now obsolete definitions that included debris in glacial transport.

Supraglacial moraines include lateral moraines (Small 1983), medial moraines (Vere and Benn 1989) and inner moraines (Weertman 1961). Lateral moraines occur at the edge of valley glaciers and comprise debris derived from the valley walls both above and below the ice. Medial moraines occur as longitudinal accumulations of debris downstream from junctions between confluent glaciers, and include debris derived from the lateral moraines of each tributary. Inner moraines are transverse accumulations of debris derived from the meltout of basal debris bands close to the glacier margin. All of these moraine types can develop into large ridges on the glacier surface as the debris cover protects the ice immediately beneath from melting while the surface of the surrounding, debris-free, ice is lowered by ablation. Thick and irregular accumulations of debris released onto the glacier surface by ablation have previously been referred to as ablation or disintegration moraine, forming part of a supraglacial land system, but these terms are increasingly being confined to terrestrial landforms that survive after ablation of the glacier. Supraglacial debris that does not form discrete topographic features on the glacier surface is not referred to as moraine.

Moraines can be formed on the ground surface subglacially or at the edge of the glacier, and by the lowering of supraglacial debris to the ground during deglaciation. They can be formed both by active (moving) ice and by stagnant ice. The principal processes by which moraines are created are the release of debris from ice by meltout and the deformation of proglacial or subglacial sediments by ice motion.

Moraines created by the lowering of supraglacial debris to the ground during deglaciation typically produce a chaotic topography and highly variable sedimentology as the landforms produced are strongly affected by resedimentation, water action and mass movement during their formation.

Subglacial moraines can occur parallel and perpendicular to ice flow or in irregular patterns. There is often a gradual transition between forms with different orientations such as Rogens (described below) and DRUMLINS. The origin of many of these features remains disputed. Areas of subglacial deposition without distinctive relief are sometimes referred to as ground moraine, but this term is falling into disuse and being replaced by non-topographic terms such as subglacial till.

Moraines parallel to ice flow include streamlined features within subglacial till, such as flutes and certain types of drumlins. The genesis of some of these features remains controversial. Whereas traditional analyses attribute them to subglacial deposition and the sculpting of subglacial deposits by moving ice, other interpretations based on subglacial meltwater processes (e.g. Shaw *et al.* 1989) imply that these features are not true moraines at all.

Transverse subglacial moraines include similar features in different locations that have been given various names and interpretations. The labels Rogen, De Geer, ribbed, washboard, corrugated, cyclic and cross-valley moraines have been applied to transverse features associated with subglacial processes. Rogen moraines are large ridges several tens of metres high, over 1km long and with crests several hundred metres apart, giving an irregular ribbed appearance to large areas of the landscape. They are often associated with flutes and drumlins, and are most commonly attributed either to thrusting of debris-rich basal ice into localized stacks beneath ice in compressive flow, or to deformation of subglacial sediment. The deformation hypothesis places Rogens at one end of a continuum of deformational forms that grades at the other end into longitudinal ridges such as flutes and drumlins. De Geer moraines are generally smaller in scale, and characterized by water-lain deposits within the moraine suggesting an origin beneath ice grounded in water.

Subglacial moraines lacking consistent orientation have been referred to as hummocky ground moraines. These are attributed either to the lowering of supraglacially released ablation moraines or to the release of debris beneath stagnant ice. The subglacial hypothesis places hummocky moraine at one end of a spectrum of forms that incorporates washboard moraines (weakly oriented hummocks) and drumlins (streamlined hummocks reflecting deposition beneath moving ice) (Eyles *et al.* 1999). Some areas of hummocky moraine have recently been reinterpreted as complex assemblages of cross-cutting and discontinuous subglacial, supraglacial and ice-marginal moraine ridges.

Ice marginal moraines occur around the edges of glaciers and are defined by their position as either lateral or frontal moraines. Moraines marking the maximum extent of a glacial advance are referred to as terminal moraines. A terminal frontal moraine is called an end moraine. Moraines deposited at successive positions of the margin during a period of progressive retreat are referred to as recessional moraines. Moraines deposited at successive positions of the margin during periods of advance are usually destroyed by the advancing ice and are not preserved in the landscape, except for the terminal moraine.

Marginal moraines at existing glaciers are typically ridges of sediment resting partially on the edge of the glacier and partially on ice-free ground beyond the margin. Upon deglaciation, ice-cored moraines lose their ice support and therefore tend to shrink in size and may become structurally unstable (Bennett *et al.* 2000). Marginal moraines may be several tens of metres in height, tens or hundreds of metres across, and may stretch for hundreds of kilometres around the margins of large ice sheets. The main processes for the formation of marginal moraines are dumping of supraglacial, englacial or basal debris transported through the glacier, and pushing of sediments previously deposited in front of the glacier.

Small push moraines can be formed by seasonal bulldozing of proglacial sediment where an ice margin oscillates with seasonally varying ablation. Larger push moraines can form by the superposition of several seasonal moraines or by a substantial advance of the margin into deformable materials. Other glacitectonic features include moraines formed by the squeezing out of deformable sediment such as saturated till from beneath the ice margin.

Meltout or dump moraines occur where englacial or supraglacial sediment is transported to the margin and dumped where the glacier ends. Dump moraines grow in size for as long as an ice margin remains *in situ* to supply sediment, their rate of growth depending on the rates of sediment supply and ablation.

The morphology and sedimentology of moraines can be used to reconstruct the characteristics of

former glaciers. The distribution of moraines reflects the geography of former glaciers and glacial process environments. Dated terminal and recessional moraines reveal the history of decay of a glacier, and process-controlled moraines reveal the locations of specific process. For example subglacial crevasse-fill ridges, which are formed by the squeezing of subglacial sediment into crevasses in the base of a glacier have been cited as indicators of glacier surging (Sharp 1985). Sediment characteristics reflect the source location of the debris: supraglacial debris is characteristically angular, while basally derived debris is typically basal faceted, subrounded and striated. Knight *et al.* (2000) showed how the distribution of clay-sized particles in a moraine reflected the distribution of a particular type of debris within the glacier that only occurred in certain process environments. Complex structures within moraines can reveal seasonal and long-term variations in processes of sedimentation. Small *et al.* (1984) showed how lateral moraine ridges derived aspects of their internal structure from seasonal variations in debris supply.

Moraines are one stage in the glacier sediment transfer system, providing long-term storage and a supply of debris to the proglacial zone. Sediment flux within glaciated basins is very sensitive to the position of glaciers relative to their moraines. When glaciers lie behind marginal moraines the bulk of sediment produced at the margin can go into storage in the moraine belt and not reach the proglacial region. When glaciers have no marginal moraines, sediment passes directly into the proglacial system. When glaciers re-advance over ancient moraines, large amounts of sediment from the moraine can be released from storage and transported into the proglacial landscape. Moraines can also focus meltwater discharge, localizing fluvial processes and causing meltwater from the glacier to be ponded up to form moraine-dammed lakes. These lakes are potentially unstable and pose a serious threat of catastrophic flooding.

Moraines are significant features within glaciated landscapes, useful indicators of past glacial activity and important components of the glacial sediment transfer system.

References

Bennett, M.R., Hambrey, M.J., Huddart, D. and Glasser, N.F. (2000) Resedimentation of debris on an ice-cored lateral moraine in the high-Arctic (Kongsvegen, Svalbard) *Geomorphology* 35, 21–40.

Dreimanis, A. (1989) Tills: their genetic terminology and classification, in R.P. Goldthwaite and C.L. Matsch (eds) *Genetic Classification of Glacigenic Deposits*, 17–83, Rotterdam: A.A. Balkema.
Eyles, N., Boyce, J.I. and Barendregt, R.W. (1999) Hummocky moraine: sedimentary record of stagnant Laurentide Ice Sheet lobes resting on soft beds, *Sedimentary Geology* 123, 163–174.
Knight, P.G., Patterson, C.J., Waller, R.I., Jones, A.P. and Robinson, Z.P. (2000) Preservation of basal-ice sediment texture in ice sheet moraines, *Quaternary Science Reviews* 19, 1,255–1,258.
Sharp, M. (1985) Crevasse-fill ridges – a landform type characteristic of surging glaciers? *Geografiska Annaler* 67A, 213–220.
Shaw, J., Kvill, D. and Rains, B. (1989) Drumlins and catastrophic subglacial floods, *Sedimentary Geology* 62, 177–202.
Small, R.J. (1983) Lateral moraines of Glacier de Tsidjiore Nouve: form, development and implications, *Journal of Glaciology* 29, 250–259.
Small, R.J., Beecroft, I.R. and Stirling, D.M. (1984) Rates of deposition on lateral moraine embankments, Glacier de Tsidjiore Nouve, Valais, Switzerland, *Journal of Glaciology* 30, 275–281.
Vere, D.M. and Benn, D.I. (1989) Structure and debris characteristics of medial moraines in Jotunheimen, Norway: implications for moraine classification, *Journal of Glaciology* 35, 276–280.
Weertman, J. (1961) Mechanism for the formation of inner moraines found near the edge of cold ice caps and ice sheets, *Journal of Glaciology* 3, 965–978.

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*, Chichester: Wiley.
Goldthwaite, R.P. and Matsch, C.L. (eds) (1989) *Genetic Classification of Glacigenic Deposits*, Rotterdam: A.A. Balkema.
Hambrey, M.J. (1994) *Glacial Environments*, London: UCL Press.
Knight, P.G. (1999) *Glaciers*, Cheltenham: Nelson Thornes.

SEE ALSO: glacial deposition

PETER G. KNIGHT

MORPHOGENETIC REGION

A morphogenetic region is an area where landforms are, or have been shaped, by the same or similar processes, mainly those controlled by climate. In climatic geomorphology there are two spatial categories: in morphoclimatic zones typical processes are considered, whereas in climato-morphogenetic regions the distinctive morphogenesis of an area is investigated. These definitions are more or less followed in continental Europe. In Anglo-American geomorphology, on the other hand, the term morphogenetic is used differently as 'the extent to which different climatic regimes are potentially capable of exerting direct and indirect influences on geomorphic processes, and thereby of generating different "morphogenetic" landform assemblages' (Chorley *et al.* 1984: 466). This nearly corresponds in German terminology to 'klimamorphologische Zonen' (morphoclimatic zones), and in French to 'les zones morphoclimatiques'. In the English literature these German and French terms are sometimes wrongly translated as 'climato-morphogenetic regions'.

The terms 'arid, humid, and nival' were introduced in 1909 by A. Penck as names for zones with distinct climate, hydrology and geomorphology. He had already recognized that these zones had shifted during the warm and cold periods of the Pleistocene and in 1913 introduced the term 'pluvial'. In 1926, in a symposion at Düsseldorf on the 'Morphologie der Klimazonen' (morphology of climatic zones), nine geomorphologists gave an overview of their research in certain areas ranging from the arctic to the humid tropics. Each one compared his findings with central Europe to stress the peculiarities. In 1948 Büdel introduced 'Das System der klimatischen Geomorphologie' (the system of climatic geomorphology). He gave a description of the typical processes in each morphoclimatic zone. The most important aspect was the interrelation of the processes in one zone, e.g. the work of a river is dependent on the relief of the area, which next to precipitation controls the amount and time of discharge. The load which has to be transported is generated from slopes and small creeks. By their interrelationships the relative strength or influence of the processes shaping the landforms should become clear. The relation between fluvial erosion and denudation was especially weighted. Thus there was some estimation of erosion rates, too. Not only were the most spectacular landforms looked for, but also the most widely distributed ones. Not only the catastrophic events but also the slowly working processes were investigated. For each morphoclimatic zone the processes were recorded as they were observed from recent occurances, from the observation of the REGOLITH, and from a check with the landforms, whose shape was interpolated and

extrapolated with the processes, a feedback. The concept of morphoclimatic zones is a very open one and may be varied, e.g. according to rock resistance (petrovariance) or tectonics (tectovariance). This is a rather broad approach and of course uncertainty or even mistakes are possible. This does not spoil the concept though. Comparison of similar regions and results from different research has increased our knowledge of the different morphoclimatic zones, though no new complete version has been made since the handbook of Büdel (1977, 1982). However, many detailed studies are founded on this concept.

For the interrelation of forming processes the terms 'Prozessgefüge' (process fabric) or 'Formungsmechanismus' (relief forming mechanism) came into use. For one of the morphoclimatic zones, the humid mid-latitudes 'zone of Holocene retarded valley building', Büdel (1982: 14) named the following components, which make up or control the 'forming mechanisms for the highly complex phenomena and processes: solution, mechanical weathering, chemical weathering, plant cover, soil development, surface denudation, linear erosion, transport, and deposition'. They are connected on 'highly complex integration levels'. In quoting 'highly complex' twice and adding 'occuring only in nature, not reproducable', he wanted to stress that on this level field measurements and laboratory experiments should be combined with the 'predominant qualitative relief analysis'. The main methods are field observations in 'natural test sites', where the phenomena are typical and which have to be searched for. Then follows the comparison with similar areas, where e.g. the influence of different rocks can be observed. Thus by comparison the petrovariance and the tectovariance can be abstracted and the processes controlled by climate become clear.

It is easier to link relief forming mechanisms to ecological factors for which climate is an abbreviation, as these comply to a zonal order, than to build a system on lithology. Of course there are distinct landforms in limestones, sandstones and granites and excellent relevant handbooks, but there is no systematic arrangement of forms due to differences in rock hardness or structure. Thus a morphogenetic region according to one of these rock groups would more or less coincide with a geological map. That would not be a new insight. It is possible to outline morphotectonic domains, but the connection to geomorphological processes is only very slowly developing, as detailed knowledge of the influence of tectonic movements on processes, except landsliding, is very small so far, and for cratons almost unknown. In both cases palaeoforms are hard to incorporate systematically, but this is easy in climatic geomorphology.

A morphoclimatic zone defined by relief forming mechanisms is a framework for detailed studies. These may be of megaforms, mesoforms or microforms and it is possible to apply many different methods. For instance, if landform facets are linked to the thickness and texture of the regolith and/or sediments, their relative age and evolution is investigated. This can be verified by laboratory research of the material, and by absolute datings. If the extent of the landforms is mapped or their changes are derived from sequences or monitoring, there is an estimation of the volume of transport possible. As this holds mainly for several hundreds or thousands of years, this provides a long-term check for short-term measurements of material transport. Thus it is possible to discriminate between natural and human-induced erosion rates. The concept of morphoclimatic zones is helpful to provide working hypotheses with regard to the full breadth of processes possible, their interdependence and relative strength. Especially in extrapolation of measured properties, an appraisal of relief forming mechanisms should be incorporated.

The essence of the concept of morphoclimatic zones is the interrelation of processes and there are almost no attempts in climatic geomorphology, as understood in Europe, to link landforms to climatic data. As in a morphoclimatic zone the interrelation of weathering, denudation and fluvial erosion is described, and it is obvious that only a general combination with climate is possible. Büdel (1977, 1982) himself delineated ten morphoclimatic zones. Originally (1948) there were twelve, and in 1963 they were reduced to five with an emphasis on the tropical semi-humid zone of excessive planation and the subpolar zone of excessive valley cutting. The names were changed too, though only slightly. This may show that zonation was not Büdel's foremost interest. He never tried to link the boundaries of the zones to climatic data. He rather insisted on the complexity of relief analysis, covering as many ecological factors as possible. There is one inconsistency, too. The zone of excessive planation should be shifted to the perhumid tropics, as only there is weathering intense enough for the concept of double planation, which is still very valid. The term climatic geomorphology is a misnomer, but the attempt to change to dynamic geomorphology was not successful as the term was introduced for a long time in contrast to tectonic geomorphology.

The difference of the Anglo-American approach to morphogenetic regions is twofold: the relevance of climatic data at the start and the broadness of the approach. The first attempt at delineating morphogenetic regions in the USA was the diagram by Peltier (1950). It was much cited but had little influence on detailed studies. Even the more sophisticated diagram of Chorley *et al.* (1984) has not been filled by regional or areal studies. Thus climatic regions as a starting point and the deduction of possible processes does not seem to be very fruitful. Instead there are single features like drainage densities connected with climatic data, or gradients of rivers or slopes are linked to sediment transport and rainfall variables. Polygenetic landforms are quite often approached from the knowledge of palaeoclimates. On the other hand there are excellent books on tropical, desert, periglacial, glacial geomorphology and karst, which describe and explain landforms and processes. But there are few interrelations and almost no connection to climatic data, though these handbooks often contain a chapter on the climate of the zone.

An extension of the morphogenetic regions was done by Brunsden in creating tectono-climatic regions. He proposed linking geotectonic domaines with morphoclimatic zones. For example, he entered into a map of the present conditions of the Indo-Australian plate the recent morphoclimatic zones and second the environmental conditions of 18,000 BP. Comparison of these two pictures gives areas of tectono-climatic stability. These are interesting hypotheses, but here, too, the starting point is the concept from facts outside geomorphology. Only later shall it be filled with field observations. It is a way that proceeds from the top downward, not from the base upwards.

It is always possible to concentrate on a special process but this should not be done in an isolated way but in the realm of the relief forming mechanism. Thus it is tied up in an analysis of interrelations of larger to smaller forms, of single processes to the process fabric. Thus the extrapolation of single processes and the interpretation of landforms becomes more secure. An example might be river terraces in mid-latitudes. Are they of climatic or tectonic origin? Not only the material of the terraces and their gradient is indicative but the origin of the pebbles and the mode of transport from the source area on a slope (e.g. by solifluction into the rivers). Such features as periglacial ice wedges casts and covers like loess are studied in relation to former climatic conditions and age. Are similar terraces developed in neighbouring areas? Which forms are incised in the older terraces? This for instance led to a detailed history of incision for the middle Rhine valley. This part of the valley is antecedent and developed during slow uplift, but the forms in detail are climate controlled. This is an example for a morphogenetic region in German understanding. There are similar regional studies in the English literature. The methods are more detailed in CLIMATO-GENETIC GEOMORPHOLOGY.

References

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

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

Chorley, R., Schumm, S.A. and Sugden, D.E. (1984) *Geomorphology*, London: Methuen.

Peltier, L.C. (1950) The geographical cycle in periglacial regions as it is related to climatic geomorphology, *Annals of the Association of American Geographers* 40, 214–236.

HANNA BREMER

MORPHOMETRIC PROPERTIES

Morphometric properties of a DRAINAGE BASIN are quantitative attributes of the landscape that are derived from the terrain or elevation surface and drainage network within a drainage basin. GEOMORPHOMETRY is the measurement and analysis of morphometric properties. Traditionally morphometric properties were determined from topographic maps using manual methods, but with the advent of geographic information system (GIS) technology, many morphometric properties can be automatically computed.

Size properties

Size variables provide measures of scale that can be used to compare the magnitudes of two or more drainage basins. Size variables are derived from measurements of the basin outline as defined by the drainage divide or are obtained from the drainage network. Many size variables are strongly correlated with one another so can be used interchangeably.

Drainage area, the two-dimensional projection of area measured in the map plane, is the most important size measure and is specified as the area contained within the drainage divide. RUNOFF GENERATION and the frequency of FLOODS is directly correlated with drainage area in many environments.

Basin length indicates the distance from the basin outlet to a point on the drainage divide, but many different methods for measuring basin length have been devised. For example, the endpoint of the length measure can be the highest point on the divide or the point on the divide that is equidistant from the outlet along the divide. Perimeter is a measure of distance around the drainage basin measured along the drainage divide.

Main channel length is the length from outlet to channel head along a subjectively defined main channel, or, more objectively, the length of the longest flow path to the drainage divide. Total channel length is the sum of lengths of all channels in a basin.

Stream order can also be used to indicate basin size (see STREAM ORDERING). The order of a basin is the order of its outlet stream. Stream magnitude is the number of FIRST-ORDER STREAMS in a basin. Magnitude is a more discerning measure of size than is order.

Surface properties

Surface properties are quantities depicted by fields comprising a value at each point within a domain (drainage basin). GIS technology provides the capability to derive surface properties from a DIGITAL ELEVATION MODEL (DEM) which is the numerical representation of an elevation surface. The elevation surface is the most fundamental surface property field, and quantifies the ground surface elevation at each point (neglecting cave and overhang special cases). DEM types include square or rectangular digital elevation grids, triangular irregular networks, sets of digital line graph contours or random points (Wilson and Gallant 2000).

The flow direction field is the direction that water flows over a surface under the action of gravity. This may be defined by the horizontal component of the surface normal. The flow direction field is represented numerically by a flow direction grid. The simplest flow direction grid is the D8 flow direction grid in which flow direction is represented by one of eight values. The value depends on which of the eight neighbouring cells (four on the main axes, four on the diagonals) is in the direction of steepest descent and thus receives its drainage. Other numerical flow direction fields can be derived using finite difference or local polynomial or surface fits to elevations of grid cells in the neighbourhood of each point (Tarboton 1997).

Terrain slope is a field giving the slope of the terrain in the direction of the flow direction field at each point. This is evaluated numerically by taking elevation differences from the elevation field over a short distance centred on each point.

Contributing area is a field representing drainage area upslope of each point. It is defined by tracing flow paths up slope from each point along the flow direction field to the drainage divide and measuring the area enclosed. Within a grid-based GIS, contributing area is evaluated by counting the number of grid cells draining to each grid cell. Contributing area is also referred to as catchment area or flow accumulation area.

Specific catchment area is a field representing contributing area per unit contour length. On a smooth surface, the contributing area to a point may be a line that has zero area. Specific catchment area is quantified using the measurable area contributing to a small length of contour (Moore *et al.* 1991: 12). Specific catchment area has units of length. On a planar surface with parallel flow, specific catchment area is equal to the upslope distance to the drainage divide.

Shape properties

Drainage basin shape is a difficult morphometric property to characterize simply, and there have been numerous attempts at defining shape variables. The simplest shape measures employ area, length, width or perimeter of the drainage basin or of a shape with area equivalent to that of the basin. More complex functions of drainage basin or drainage network shape are best portrayed using two-dimensional graphical plots.

The cumulative area distribution function is defined as the proportion of a drainage basin that has a drainage area greater than or equal to a specified area. It is typically represented by plotting cumulative area versus area on a log-log line chart.

The distance area diagram depicts the area of the basin as a function of distance along flow paths to the outlet. The channel network width function is the number of channels at a given distance from the drainage basin outlet, as measured along the drainage network, and is typically plotted as a line or bar chart. The distance area diagram and channel network width function both give an indication of basin hydrological response and are related to the instantaneous unit hydrograph.

Relief properties

RELIEF properties bring the dimension of height into morphometric analysis. Because many landscape processes are driven by gravity, relief properties are frequently used as indicators of EROSION potential and DENUDATION rates.

Total basin relief is the difference in height between the outlet and the highest point on the drainage divide. Relief ratio removes the size effect by dividing total relief by basin length. Sediment yield (see SEDIMENT LOAD AND YIELD) in small drainage basins has been shown to be exponentially related to relief ratio (Hadley and Schumm 1961: 172).

A more complex representation of basin relief is the area-elevation relationship or hypsometric curve. The hypsometric curve is a plot of the area of a basin (on the x-axis) above each elevation value (on the y-axis). The axes are commonly normalized to range between zero and one. The hypsometric curve is equivalent to one minus the cumulative distribution of elevation within a drainage basin. Davisian model evolutionary stage can be inferred from the shape of a basin's hypsometric curve.

Texture properties

Texture indicates the amount of landscape dissection by a channel network. The contours on a map of a highly textured landscape will have many small crenulations (wiggles) indicating the presence of numerous channels.

DRAINAGE DENSITY (Horton 1945: 283), the best-known texture indicator, is defined as the lengths of all stream channels in a drainage basin divided by drainage area and has units of 1/length. Drainage density ranges from less than $1\,km^{-1}$ to over $800\,km^{-1}$, attaining maximum values in semi-arid areas (Gregory 1976: 291). High drainage densities indicate highly textured landscapes, short hillslopes and domination by OVERLAND FLOW runoff typical of BADLANDS.

The area–slope relationship quantifies the area draining through a point versus the terrain slope at that point, typically plotted on a log-log scale graph. The scatter when all points or grid cells are used is removed by binning (e.g. using a moving average) to reveal a characteristic area–slope relationship with two distinct regions. For small areas, slope increases with drainage area and for large areas, slope decreases with area. The turnover point in the relationship has been interpreted as the drainage area at which diffusive hillslope processes (see HILLSLOPE, PROCESS) are overtaken by fluvial processes and channels are initiated (Tarboton *et al.* 1992: 73).

References

Gregory, K.J. (1976) Drainage networks and climate, in E. Derbyshire (ed.) *Geomorphology and Climate*, 289–315, London: Wiley.

Hadley, R.F. and Schumm, S.A. (1961) *Sediment Sources and Drainage-Basin Characteristics in the Upper Cheyenne River Basin*, Washington: US Geological Survey Water Supply Paper 1,531.

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.

Moore, I.D., Grayson, R.B. and Ladson, A.R. (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, *Hydrological Processes* 5, 3–30.

Tarboton, D.G. (1997) A new method for the determination of flow directions and contributing areas in grid digital elevation models, *Water Resources Research* 33, 309–319.

Tarboton, D.G., Bra, R.L. and Rodriguez-Iturbe, I. (1992) A physical basis for drainage density, *Geomorphology* 5, 59–76.

Wilson, J.P. and Gallant, J.C. (2000) *Terrain Analysis: Principles and Applications*, New York: Wiley.

Further reading

Gardiner, V. (1975) *Drainage Basin Morphometry*, British Geomorphological Research Group Technical Bulletin No. 14, Norwich: GeoAbstracts.

SEE ALSO: Horton's Laws

CRAIG N. GOODWIN AND DAVID G. TARBOTON

MORPHOTECTONICS

Morphotectonics is the term pertinent to links between geomorphology and tectonics, although individual authors apparently understand the exact nature of these links in slightly different ways. Most often, morphotectonics is considered synonymous with TECTONIC GEOMORPHOLOGY and defined simply as the study of the interaction of tectonics and geomorphology. Embleton (1987) lists four main lines of interest in morphotectonic research: (1) study of landforms indicative of contemporary or recent tectonic movement, (2) study of deformation of PLANATION SURFACES, (3) study of geomorphological effects of earthquakes (see SEISMOTECTONIC GEOMORPHOLOGY), (4) use of geomorphological evidence to predict earthquakes. It needs to be emphasized that in some countries morphotectonics is a term of very limited usage. For example, two recent American textbooks about tectonic geomorphology (Burbank and Anderson 2001; Keller and Pinter 2002) do not mention morphotectonics, although they evidently deal with this kind of phenomenon.

Fairbridge (1968) offers a different explanation and understands morphotectonics as a means to classify major landforms of the globe rather than any landforms related to tectonic processes. Accordingly, he distinguishes morphotectonic units of first and second order. In the first order these are continents and oceanic basins, in the second one there are shields, younger mountain belts, older mountain massifs, basin-and-range areas, rift zones and basins. This global context of morphotectonics is also evident in the study of great ESCARPMENTS (Ollier 1985).

In practice, the morphotectonic approach frequently means using landforms or any other surface features (e.g. drainage patterns) as a key to infer the existence of tectonic features, especially in relatively stable areas where seismicity and present-day rates of uplift and subsidence are negligible. They acquire the status of geomorphic markers of tectonics. Geomorphological maps, drainage pattern maps, digital elevation models and their various derivatives are analysed with the aim of locating anomalies in landform distribution, river courses, channel form, terrace profiles, local relief or specific landforms such as slope breaks. These anomalies in turn, if no other explanation for their occurrence is available, are considered to reflect the presence of tectonically active zones or areas. Detailed analysis of river patterns can be a particularly valuable tool in morphotectonic research in lowland areas, where hardly any other evidence is at hand.

References

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

Embleton, C. (1987) Neotectonic and morphotectonic research, *Zeitschrift für Geomorphologie N.F., Supplementband* 63, 1–7.

Fairbridge, R.W. (1968) Morphotectonics, in R.W. Fairbridge (ed.) *Encyclopedia of Geomorphology*, 733–736, New York: Reinhold.

Keller, E.A. and Pinter, N. (2002) *Active Tectonics*, Englewood Cliffs, NJ: Prentice Hall.

Ollier, C.D. (ed.) (1985) Morphotectonics of passive continental margins, *Zeitschrift für Geomorphologie N.F., Supplementband* 54.

Further reading

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

Ollier, C.D. (1981) *Tectonics and Landforms*, London: Longman.

Schumm, S.A., Dumont, J.F. and Holbrook, J.M. (2000) *Active Tectonics and Alluvial Rivers*, Cambridge: Cambridge University Press.

Summerfield, M.A. (ed.) (2000) *Geomorphology and Global Tectonics*, Chichester: Wiley.

SEE ALSO: active and capable fault; active margin; fault and fault scarp; global geomorphology; neotectonics; passive margin

PIOTR MIGOŃ

MOULIN

Moulins, or glacier mills, are sink holes that owe their name to the roaring noise of water that engulfs itself in them. They form in the ablation zone of GLACIERS (Paterson 1994), where meltwater (see MELTWATER AND MELTWATER CHANNEL) manages to cut stream channels into the ice, generally parallel to glacier slope. These channels are eventually intercepted by crevasses, which form perpendicular or oblique to glacier slope in response to ICE flow related to bedrock surface irregularities. Moulins are the result of meltwater flowing into crevasses (Rothlisberger and Lang 1987).

Moulins are characterized by a vertical shaft up to 100 m tall, developing along the planes of single or cross-cutting crevasses and prolonging into a downflow dipping gallery that follows structures related to glacial flow. The gallery dips approximately 45° and forms a succession of pools on an irregular floor, but the gallery sometimes dips approximately parallel to glacier slope when the shaft is less than 50 m tall. Shafts are circular or elliptical in horizontal cross section and range between less than 1 m to over 20 m in their long axis, but detail of their morphology is controlled by the dynamics of the water that flows into it (Holmlund 1988).

The bottom of moulins is often submerged. Water level can vary within a few hours and from one season to the next in relation to meteorological conditions, glacial flow, ice plasticity and according to the facility with which water can flow along the base of the glacier.

During the summer, moulins provide the main inputs of glacial aquifers. During the winter, they are separated from the surface by a snow bridge. However, water level in the moulins usually increases during the winter and then decreases in a jerky fashion, the moulins functioning like surge tanks (Schroeder 1998). This implies that drainage in the glacier tends to clog in the front first during the beginning of the cold season, while the water column that then remains stocked within the moulins prevents glacial flow from closing it. With the onset of the warm season, this water that was stocked within the moulins helps in reinitiating subglacial drainage.

The life expectancy of moulins can reach several dozens of years. Moving along with the glacier, they eventually lose their connection to glacier surface drainage at the profit of new moulins forming upflow from them. In dead ice, moulins often reach down through the entire glacier. Megapotholes (diameter > 50 m) developed in rock bars of regions that were glaciated during the Quaternary are thought to be the result of extended water circulation at the base of moulins in stagnant ice.

References

Holmlund, P. (1988) Internal geometry and evolution of moulins, Storglaciären, Sweden, *Journal of Glaciology* 34, 242–248.

Paterson, W.S.B. (1994) *The Physics of Glaciers*, New York: Pergamon.

Rothlisberger, H. and Lang, H. (1987) Glacial hydrology, in A.M. Gurnell and M.J. Clark (eds) *Glaciofluvial Sediment Transfer*, 209–284, New York: Wiley.

Schroeder, J. (1998) Hans glacier moulins observed from 1988 to 1992, Svalbard, *Norsk Geografisk Tidsskrift* 52, 79–88.

JACQUES SCHROEDER

MOUND SPRING

Small mounds formed preferentially along fault lines by artesian springs. Solutes and colloids are precipitated to form travertines or tufas (see TUFA AND TRAVERTINE) of calcium carbonate, together with various siliceous and ferruginous deposits. Wind-blown sand and accumulated plant debris, together with mud and sand carried up with the spring water, assist in their formation. Where springs display high rates of water flow there tends to be little or no mound formation – they are too erosive. However, springs with low discharge rates and laminar flow experience high rates of evaporation (especially in arid environments) and have a greater possibility of accumulating chemical precipitates.

Major examples of such features are known from the Great Artesian Basin of Central Australia (Ponder 1986) and from the depressions of the Western Desert in Egypt, where much accretion has occurred when vegetated fields, irrigated by the springs, have trapped aeolian sediment (Brookes 1989).

References

Brookes, I.A. (1989) Above the salt: sediment accretion and irrigation agriculture in an Egyptian oasis, *Journal of Arid Environments* 17, 335–348.

Ponder, W.F. (1986) Mound springs of the Great Artesian Basin, in P. de Deckker and W.D. Williams (eds) *Limnology in Australia*, 403–420, Dordrecht: Junk.

A.S. GOUDIE

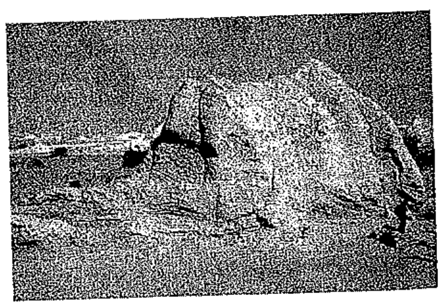

Plate 82 A silty mound deposit associated with spring activity in the Farafra Oasis of the Western Desert of Egypt

MOUNTAIN GEOMORPHOLOGY

Mountain geomorphology is a 'regional component within geomorphology' (Barsch and Caine 1984). The region in this case is the world's mountains defined by absolute elevation (> 600 m above sea level), available relief (> 200 m km^{-2}) and topographic slopes (> 10°). There is no international standard definition, but other elements which are frequently incorporated are high spatial variability, presence of ice and snow and evidence of late Pleistocene glaciation. Carl Troll (1973) who was the modern creator of mountain geomorphology, defined mountain systems as those which encompass more than one vegetation belt, but do not reach alpine elevation by contrast with high mountain systems (hochgebirge) which extend above the timberline.

Fairbridge (1968) rehearses the classification of mountains by scale and continuity: (a) mountain is a singular, isolated feature or a feature outstanding within a mountain mass; (b) a mountain range is a linear topographic feature of high relief, usually in the form of a single ridge; (c) a mountain chain is a term applied to linear topographic features of high relief, but usually given to major features that persist for thousands of kilometres; (d) a mountain mass, massif, block or group is a term applied to irregular regions of mountain terrain, not characterized by simple linear trends; and (e) a mountain system is reserved for the greatest continent-spanning features.

A simple genetic system of mountain types, which was developed before global plate tectonics was understood, is nevertheless useful in local-scale understanding. There are two broad categories: (1) structural, tectonic or constructional forms and (2) denudational or destructional forms. Under the first category can be identified: (a) volcanic; (b) fold and nappe; (c) block; (d) dome; (e) erosional uplift or outlier; (f) structural outlier or klippe; (g) polycyclic tectonic; and (h) epigene mountains. Under the second category are defined: (a) differential erosion; (b) exhumed; (c) plutonic and metamorphic complex; and (d) polycyclic denudational mountains (Fairbridge 1968). In the structural mountain categories it is the tectonic process that has played a primary role; in the denudational categories it is the denudational processes that are primary. The lithology and the climatic history are both extremely important with respect to the detailed modification of these mountain types. Indeed, much of the science of geomorphology is centred on the discrimination of these second-order effects.

The simplest typology of mountain geomorphology makes use of the tripartite division into historical, functional and applied mountain geomorphology. Historical mountain geomorphology focuses on the evolution of mountains and mountain systems over both long and medium timescales. It is common, at the largest scale, to distinguish between young active mountain belts which have evolved throughout the Cenozoic and are still associated with active plate margins and mountains on passive continental margins. Nearly all the literature on mountain building in the past forty years has concentrated on active margins where collision and subduction may explain both mountains and the structures within them. Most exciting in recent years is the trend towards quantifying rates of uplift and denudation with the development of new geochemical, geochronological and geodetic methods. But, in reality, there are mountains on PASSIVE MARGINS too (Ollier and Pain 2000). The evolution of these older mountain belts is intrinsically more complex as they do not easily fit into the simple plate tectonic story of mountain building at collisional sites and include the history of the Earth since the breakup of Gondwanaland during the Mesozoic. A major difference of opinion has emerged between those who place greatest emphasis on the data from FISSION TRACK ANALYSIS and those who use whatever landform, stratigraphic and geological data that can be found to constrain the interpretation. Whereas geomorphic models of denudation history are difficult to validate, interpretation of fission track data in terms of denudation history is complex.

Functional geomorphology of mountains includes the assessment of processes, rates and spatial and temporal patterns of mountain belt erosion. The process framework should ideally involve a consideration of the coupling of uplift and erosion; many geomorphic models have failed to include realistic models of this coupling. In mountain belts, such a consideration is obligatory as both uplift rates and erosion rates achieve maximum values and the coupling of the processes is even more critical than in lowland regions. Improvements in understanding of fluvial bedrock incision processes, hillslope mass wasting, glacial valley lowering and SEDIMENT ROUTING are leading to the development of improved mountain landscape evolution models.

Feedbacks between tectonic, climatic and geomorphic processes have been explored in geodynamic models and solid, solute and organic fluxes from mountain belts have been constrained and considered within a global geochemical context. The topographic evolution of mountain belts can be modelled with increasing realism, but the issue of equilibrium conditions versus transience is still far from resolved.

APPLIED GEOMORPHOLOGY of mountains: mountain habitats create or magnify natural hazards in the form of dangerous geomorphic processes. The interaction of geomorphic processes with mountain societies, their land uses and their response capabilities determines risk. Recent social and environmental changes in the mountains has led to the modernization of the natural hazard problematique. As a result, planned responses, including mitigation strategies for specific hazards and mountain disasters, must be developed to reduce the vulnerability of mountain peoples. The applied mountain geomorphologist has a distinctive role to play.

There are three formal or semi-formal attempts to define the field in the literature: Hewitt (1972) Barsch and Caine (1984) and Slaymaker (1991).

Hewitt addressed two issues: the idea of a high energy condition and the relation of distinctive morphological features to clima-geomorphic conditions and denudation history. These he said express the distinctiveness of mountain geomorphology. Under the high energy condition, he dealt with regional rates of net erosion, magnitude and frequency of erosional events and energy in the mountain geomorphic system. Under distinctive morphological features, he picked out accordant erosion surfaces, valley asymmetry and threshold slopes for detailed treatment. The beauty of Hewitt's vigorous statement is that he foreshadows the increasingly heavy emphasis on the operation of geomorphic processes in mountain regions, but also warns of the danger of not relating these observations to the larger questions of mountain landscapes and neglecting to either solve them or to restate them in better terms.

Barsch and Caine (1984) divide mountain geomorphology into studies of mountain form and morphodynamics in mountains. These two categories they further subdivide into (a) morphometry and structure, (b) relief generation and history, (c) morphoclimatic models and (d) process dynamics and activity. Morphometry and structure depend heavily on plate tectonic setting of the mountains. There are four convergent plate settings in which some of the most rapidly evolving mountain systems of the world are located. These are: oceanic to oceanic plate convergence (e.g. Japanese Alps and the Aleutian Arc, Alaska); oceanic to continental plate convergence (e.g. South Island, New Zealand and Cascade Ranges, Pacific North-West); continental to continental plate convergence (e.g. Himalayas); and displaced terranes along accreted margins (e.g. British Columbia). Divergent plate settings include sites of oceanic spreading, such as Iceland and the Galapagos Islands, and intra-continental rifts, such as the Gulf of Aqaba and the Scottish Highlands. Transform plate settings are threefold: ridge past ridge (e.g. Coast Ranges of California); trench past trench (e.g. Anatolia, Turkey) and ridge past trench (e.g. Pakistan–Afghanistan). It is not difficult to understand why mountains are preferentially located in all of the above plate marginal locations. But mountains are also found in plate interior settings, such as the following: hot spots (e.g. Hawaii and Yellowstone National Park); continental flood basalts (e.g. Deccan, India and Columbia Plateau, Pacific North-West); shields (e.g. Ahaggar Mountains, Sahara); intracratonic uplift sites (e.g. the San Rafael Swell, Utah); post-tectonic magmatic intrusion sites (Air Mountains, Nigeria); and evaporite diapers (e.g. Zagros Mountains, Iran). They note that in most mountain regions, the balance between denudation and tectonic uplift is resolved in favour of the latter. They fail to differentiate between high mountains and mountain systems on the basis of morphometry alone, but they make the case that there are four distinctive 'relief types' within high mountain systems, namely the Alp type, the Rocky Mountain type, polar mountains and desert mountains. The Alp type is associated with an overriding impact of glacial ice and glacial erosion; the Rocky Mountain type has a less pervasive impact of glacial erosion and includes areas of low relief on flat summits and rounded interfluves; polar mountains give evidence of intense glaciation, but are frequently with a local relief of less than 1,000 m; and desert mountains are high mountains in the 'true' sense even though they do not reach timberline and were only lightly glaciated during the Pleistocene.

Relief development in high mountains revolves around questions of accordant surfaces and valley benches as indicators of mode of valley dissection. Attention is directed to (a) the alpine summit accordance or 'gipfelflur', which has been explained as a remnant of an old erosion surface; (b) the alpine crest and summit accordance, explained as the product of regular patterns of dissection which constrain summits to approximately the same elevation; (c) the timberline and alp slope accordance, explained as an inter-glacial alp slope associated with a higher timberline than the present one; and (d) benches along the sides of major valleys, variously explained as Tertiary, Pleistocene glacial and inter-glacial timberline effects. Ford *et al.* (1981) have suggested that the age of the present relief of the southern Canadian Rockies is Pliocene, considerably older than had previously been thought.

Building on Caine (1974), Barsch and Caine (1984) distinguish four mountain geomorphic process systems: (1) the glacial system; (2) the coarse debris system; (3) the fine clastic sediment system; and (4) the geochemical system. Of the four, the glacial and the coarse debris systems are most characteristic of high mountain terrain. The final section of their paper summarizes contemporary geomorphic activity in high mountain areas using calculations of sediment flux (in $\mathrm{J\,km^{-2}\,yr^{-1}}$) from Sweden, Switzerland and the United Sates (Rapp 1960; Jackli 1957; Caine 1976). Most interesting was the observation that talus shift, solifluction, soil creep and other processes of slow mass wasting accounted for no more than 15 per cent of the geomorphic work done in all three areas and their relative importance decreased with increasing size of basin. The authors suggest that there are two urgent needs for mountain geomorphology: (1) linking process and form in a meaningful way and (2) identifying anthropogenic influences and ways in which they may be propagated through the mountain system.

Slaymaker (1991) suggests that the meso and macro-scales are the only spatial scales at which a distinctive mountain geomorphology signal is likely to be apparent. He then adopts a slightly modified version of the Chorley and Kennedy (1971) open systems framework to identify five mountain systems: (1) morphological; (2) morphologic evolutionary; (3) cascading; (4) process-response; and (5) control systems. Each of these mountain systems is examined at meso- and macro-scales in search of characteristic mountain geomorphology forms and processes. He claims that this typology is useful in that different measurement programmes are appropriate within each of the ten mountain systems identified.

In fact, these ten mountain geomorphic systems serve to underline the huge variety of forms and processes that characterize mountain geomorphology and support the contention that mountain geomorphology is characterized by its extreme gradients, not only of topography, but also of energy and mass balances and ecological responses. High vertical and horizontal rates of change over space of landforms and processes and rapid rates of change over time distinguish mountain geomorphic systems from other regions. Hence the validation of mountain geomorphology as a regional component within geomorphology.

References

Barsch, D. and Caine, N. (1984) The nature of mountain geomorphology, *Mountain Research and Development* 4, 287–298.

Caine, N. (1974) The geomorphic processes of the alpine environment, in J.D. Ives and R.G. Barry (eds) *Arctic and Alpine Environments*, 721–748, London: Methuen.

——(1976) A uniform measure of sub-aerial erosion, *Geological Society of America Bulletin* 87, 137–140.

Chorley, R.J. and Kennedy, B.A. (1971) *Physical Geography: A Systems Approach*, London: Prentice Hall.

Fairbridge, R.W. (1968) *The Encyclopedia of Geomorphology*, New York: Reinhold.

Ford, D.C., Schwarcz, H.P., Drake, J.J., Gascoyne, M., Harmon, R.S. and Latham, A.G. (1981) Estimations of the age of the existing relief within the southern Rocky Mountains of Canada, *Arctic and Alpine Research* 13, 1–10.

Hewitt, K. (1972) The mountain environment and geomorphic processes, in O. Slaymaker and H.J. McPherson (eds) *Mountain Geomorphology*, 17–34, Vancouver: Tantalus.

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

Ollier, C.D. and Pain, C.F. (2000) *The Origin of Mountains*, London: Routledge.

Rapp, A. (1960) Recent development of mountain slopes in Karkevagge and surroundings, northern Scandinavia, *Geografiska Annaler* 42-A, 73–200.

Slaymaker, O. (1991) Mountain geomorphology: a theoretical framework for measurement programmes, *Catena* 18, 427–437.

Troll, C. (1973) High mountain belts between the polar caps and the equator: their definition and lower limit, *Arctic and Alpine Research* 5, 19–28.

SEE ALSO: plate tectonics

OLAV SLAYMAKER

MUD FLAT AND MUDDY COAST

The term mud is used to refer to sediments comprised chiefly of silts (size range 4 to 63 μm) and clays (finer than 4 μm). Such fine material is readily maintained in suspension and can be transported over long distances by coastal currents. Unlike coarser sands and gravels, muddy sediments tend to be cohesive. The electrochemical properties of clay mineral particles mean that these can bind together to form larger composite particles in a process known as flocculation. Flocculation is influenced by a variety of factors, notably salinity, fluid shear and suspended sediment concentration (Lick and Huang 1993). The effect of these processes may vary over quite short spatial and temporal scales, especially in estuaries, where mixing of freshwater and salt-water occurs and where marked variation in flow intensity occurs at tidal timescales. The cohesive nature of muddy sediments makes their behaviour far more complex than that of non-cohesive sands. Flocculated sediments typically settle from suspension far more rapidly than their constituent mineral particles, and the stability of natural muddy deposits is governed not only by physical processes but also by the activity of a rich and diverse biota including macroscopic and microscopic algae, invertebrates and bacteria (Paterson 1997).

Muddy coasts typically occur along low energy shorelines that are well supplied with silt and clay-sized sediments. They include many estuarine margins, delta shorelines, and areas of open coast subject to low wave energy. Such settings are usually dominated by tidal processes, and the characteristic landforms of muddy coasts – SALTMARSHes, MANGROVE SWAMPs and tidal flats – are often very well developed under macro-tidal conditions (Hayes 1975). Enormous quantities of muddy sediment are supplied by some of the world's major rivers, and their estuaries and deltas often feature extensive shore-attached mud banks. Open coast mud banks occur downdrift of major fluvial sediment sources, notably in the Gulf of Mexico (associated with the Mississippi River); more than 850 km of the Jiangsu coastline of China, supplied by the Huanghe and Changjiang Rivers (Ren 1987); and along the south-west coast of India. Both estuarine and open coast mud banks are highly dynamic landforms, which exhibit both seasonal and decadal style variability in response to variations in river flow and wave energy. Their deposits often have a high water content and include highly mobile 'fluid muds' that are highly effective in dissipating incident wave energy (Mehta and Kirby 2001). In other environmental settings, coastal and marine sediment sources are more important. In the North Sea, for example, erosion of unconsolidated Quaternary cliffs provides a major source of muddy sediments along the coast of eastern England (Ke *et al.* 1996).

The intertidal zone of muddy coasts typically comprises: a lower zone, characterized by sandy tidal flats; a middle zone of muddy tidal flats; and an upper intertidal of vegetated saltmarsh or mangrove. In some localities, the upper intertidal grades into a high supratidal plain or flat, inundated only by extreme water levels (e.g. during storm surges). The low topography is dominated by low gradient surfaces, crossed by shallow tidal channels (or 'creeks'). These channels vary in complexity from single 'rills' to intricate networks, and are generally best developed within the mud flat and saltmarsh sub-environments. Surface sediments generally decrease in grain size in a landward direction and the vertical stratigraphic sequence generally exhibits a fining upward sequence, usually attributable to transitions between tidal flat and saltmarsh as sedimentation proceeds.

The physical processes of mud flat sedimentation have been extensively studied, mainly from the perspective of sediment transport and deposition, with rather less emphasis being placed on processes of deposit consolidation and erosion (Amos 1995). A reduction in tidal current velocities in a landwards direction leads to the deposition of sediment suspended during the flood tide: the diminution in the competence of flows to transport material also explains the landward reduction in grain size. Although a portion of the newly deposited material is resuspended on the ebb tide, vertical and horizontal accretion of muddy intertidal sediments implies the dominance of flood-tide deposition (Evans 1965).

In the absence of any net (or 'residual') landward water transport, the accumulation of mud is further explained by reference to the concepts of 'settling lag' and 'scour lag'. Both these concepts were developed in the 1950s to account for tidal flat sedimentation in the Dutch Wadden Sea (see Amos 1995 for a recent review and evaluation of this work). Settling lag refers to the time elapsed between the slackening of tidal current intensity below the threshold of suspension for a given sediment and the deposition of the particle on the bottom. This means that particles are deposited some distance landwards of the point at which settling from suspension commences. Scour lag is a consequence of the higher flow intensity required to re-entrain a particle once it has been deposited. This is especially important for cohesive sediments, and means that ebb-directed transport occurs over a shorter duration than that of the flood tide. Both mechanisms tend to encourage the landward transport of mud and its accumulation in shallow intertidal areas.

Rates of mud flat sedimentation may be initially rapid (of the order of several centimetres a year), but diminish as the build-up of elevation reduces the frequency of inundation. Colonization by halophytic vegetation (and a transition to saltmarsh or mangrove) may be associated with a further increase in sedimentation rate owing to increased sediment retention under an energy-dissipative plant cover. However, this rapidly diminishes as vertical accretion further reduces inundation and wetland surface elevations tend towards a state of equilibrium between further sedimentation, the compaction of earlier deposits and sea level.

Mud flat topography arises from the dynamic interaction of tidal and wave-related hydrodynamics, sedimentation and morphology itself. Recent work has shown wave action to be more important than previously thought and has also drawn attention to the importance of biological processes in mediating sediment stability. Pethick (1996) draws an analogy between the morphological adjustment of mud flats to variations in wave energy and the morphodynamics of non-cohesive sandy beaches. The influence of waves differs between inner estuary sites, subject to small (fetch-limited) waves and outer estuary or open coastal sites, which experience a greater range of wave heights. At fetch-limited sites, waves may still exert oscillatory shear-stresses which exceed those generated by tidal currents and which are capable of resuspending mud flat sediments. A zone of resuspension migrates up and down the mud flat profile with the tidal variation in water level. Over time, the mud flat profile adjusts towards a form that is in equilibrium with wave induced stresses. The resulting profile is typically concave, a finding supported by numerical modelling experiments undertaken by Roberts *et al.* (2000). At more exposed sites, mud flats may undergo more episodic erosional adjustments in response to high wave energy conditions. In this case, the balance between individual erosion events and depositional recovery in the intervening periods determines longer term mud flat morphology.

Predominantly accretional mud flats tend to have a high and convex profile, whilst erosional mud flats are typified by a lower and concave profile. Mehta and Kirby (2001) attribute the contrasting stability of these mud flat morphologies to differences in their dissipative characteristics. In the case of high, convex mud flats, flexing of water-sediment mixture substantially dissipates tidal and wave-induced stresses, especially where thin surficial fluid mud layers are present. In low, concave mud flats, however, deposits are normally overconsolidated, such that hydrodynamic stresses are dissipated in overcoming interparticle cohesion, and in entraining sediment. Such systems are likely to be erosional.

The surficial sediments of mud flats support a variety of organisms, some of which act to stabilize the sediment and some of which act to increase the likelihood of erosion. Most mud flats support dense communities of benthic diatoms, which excrete large quantities of extracellular polymeric substances (EPS). EPS consist mainly of polysaccharides compounds and are a major component of surface films, which increase the stability of the sediment surface (Paterson 1997). Meso- and macro-fauna are active over a greater depth and may variously stabilize sediment (e.g. through the construction of EPS-coated tubular burrows) or reduce stability (e.g. by grazing on the micro-algae which helps bind sediment particles, or by reworking sediments through burrowing). Biological processes are extremely variable, both spatially and temporally, and are extremely important in determining the threshold stress at which erosion occurs. Once this threshold is exceeded, however, erosion may proceed more rapidly at a rate more closely controlled by bulk sediment properties.

Mud flats are increasingly valued as a habitat for large invertebrate populations which, in turn, provide a vital food source for wading birds. As landforms they are also of engineering significance as naturally dissipative systems which, allied to fixed defences, can provide an important component of integrated and more sustainable strategies for coastal protection. From both these perspectives, high and convex mud flats are preferable to low and concave ones (Kirby 2000). In the former case, waves are progressively attenuated as they approach the shore, a process which is further assisted by any saltmarsh fringe. Convex mud flats tend to have a larger invertebrate fauna, concentrated at a higher elevation within the tidal range, and capable of sustaining greater bird and fish populations. In contrast, erosional concave mud flats are less effective in dissipating wave energy and, in their upper portions, prone to rotational failure and slumping, with adverse consequences for the stability of sea defences.

References

Amos, C.L. (1995) Siliclastic tidal flats, in G.M.E. Perillo (ed.) *Geomorphology and Sedimentology of Estuaries*, 273–306, Amsterdam: Elsevier.

Evans, G. (1965) Intertidal flat sediments and their environment of deposition in The Wash, *Quarterly Journal of the Geological Society of London*, 121, 209–245.

Hayes, M.O. (1975) Morphology of sand accumulations in estuaries', in L.E. Cronin (ed.) *Estuarine Research, Volume II*, 3–22, New York: Academic Press.

Ke, X., Evans, G. and Collins, M. (1996) Hydrodynamics and sediment dynamics of The Wash embayment, eastern England, *Sedimentology* 43, 157–174.

Kirby, R. (2000) Practical implications of tidal flat shape, *Continental Shelf Research* 20, 1,061–1,077.

Lick, W. and Huang, H. (1993) Flocculation and the physical properties of flocs, in A.J. Mehta (ed.) *Nearshore and Estuarine Cohesive Sediment Transport*, 21–39, Washington, DC: American Geophysical Union.

Mehta, A.J. and Kirby, R. (2001) Muddy coast dynamics and stability, *Journal of Coastal Research, Special Issue* 27, 121–136.

Paterson, D.M. (1997) Biological mediation of sediment erodibility: ecology and physical dynamics, in N. Burt, R. Parker and J. Watts (eds) *Cohesive sediments*, 215–229, Chichester: Wiley.

Pethick, J.S. (1996) The geomorphology of mudflats, in K.F. Nordstrom and C.T. Roman (eds) *Estuarine Shores: Evolution, Environment and Human Alterations*, 185–211, Chichester: Wiley.

Ren, M. (ed.) (1987) *Modern Sedimentation in the Coastal and Nearshore Zones of China*, New York: Springer-Verlag.

Roberts, W., Le Hir, P. and Whitehouse, R.J.S. (2000) Investigation using simple mathematical models on the effect of tidal currents and waves on the profile shape of intertidal mudflats, *Continental Shelf Research* 20, 1,079–1,097.

Further reading

Healy, T., Ying Wang and Healy, J.A. (eds) (2002) *Muddy Coasts of the World: Processes, Deposits and Function*, Amsterdam: Elsevier Science.

SEE ALSO: mangrove swamp; saltmarsh; tidal creek; tidal delta

J.R. FRENCH

MUD VOLCANO

Mud volcanoes are positive topographic features formed by periodic venting of fluid mud, water and hydrocarbons (Kopf 2002). Individual mud volcanoes are elliptical mounds up to 2,000 m in diameter and 100 m in height. Cones and pools are often concentrated near the summit, and active portions of mud volcanoes are hummocky, unvegetated and covered by mud flows. Although clay and silt dominate mud-volcano deposits, pebble- to boulder-size clasts are common.

Mud volcanoes are known from approximately thirty regions worldwide. Most examples occur in compressional tectonic settings such as convergent plate margins. However, mud volcanoes also occur along passive margins and continental interiors. Excellent examples of subaerial mud volcanoes are present in Azerbaijan, Burma, Colombia, Indonesia, Iran, Italy, Mexico, Pakistan, Panama, Trinidad and Venezuela (Higgins and Saunders 1974). Subaqueous mud volcanoes occur in the Gulf of Mexico and the Barbados Ridge accretionary prism.

Mud volcanoes are commonly underlain by thick sequences of organic and clay-rich sediments (Hedberg 1974). Rapid sedimentation combined with methane generation, clay mineral diagenesis and tectonic compression produces high pore-fluid pressures, which mobilize fluid mud. Mud, water and hydrocarbons, migrate upward along fractures and faults that are typically associated with mud diapir-cored anticlines. If pressures are sufficient, fluid mud erupts at the seafloor or on the land surface to form mud volcanoes. In some instances, violent eruptions are accompanied by gas flares.

References

Hedberg, H.D. (1974) Relation of methane generation to undercompacted shales, shale diapirs, and mud volcanoes, *American Association of Petroleum Geologists Bulletin* 58, 661–673.

Higgins, G.E. and Saunders, J.B. (1974) Mud volcanoes – their nature and origin, in P. Jung (ed.) *Contributions to the Geology and Paleobiology of the Caribbean and Adjacent Areas*, 84, 101–152, Basel: Verhandlungen der Natureforschenden Gesellschaft.

Kopf, A.J. (2002) Significance of mud volcanism, *Reviews of Geophysics* 40(2), 1–52.

SEE ALSO: diapir; liquefaction; mudlump

ANDRES ASLAN

MUDLUMP

A diapiric structure composed of fine-textured material, especially clay, formed near the mouth of a delta's distributary. Mudlumps range in size from pinnacles to small, elongated islands. They are both subaqueous and subaerial with the subaerial forms often subject to rapid erosion by waves. The surface of mudlumps is usually irregular and most have gas (methane) and mud vents. Although several theories have been proposed for their formation, it is now generally accepted that they are the result of the intrusion of plastic, prodelta clay through overlying sand layers. They develop in sequence as distributary mouths advance seaward. Originally thought to have been unique to the distributaries of the Mississippi River, mudlumps are now known to exist in a few other deltaic areas.

Further reading

Lyell, C. (1889) *Principles of Geology*, Vol. 1, 11th edition, New York: Appleton.

Morgan, J.P., Coleman, J.M. and Gagliano, S.M. (1968) Mudlumps: diapiric structures in Mississippi delta sediment, in *Diapirism and Diapirs*, American Association of Petroleum Geologists, Memoir 8, 145–161.

Walker, H.J. and Grabau, W.E. (1992) Mudlumps, in D.G. Janelle (ed.) *Geographical Snapshots of North America*, 211–214, New York: Guilford.

H. JESSE WALKER